SOME NEW RESULTS ON TWO SIMPLE TIME SERIES MODELS-- PREDICTION COVERAGE FOR AR(1) AND MODEL BUILDING FOR JITTERY COSINE WAVES

BY

PAN-YU LAI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1985

ACKNOWLEDGMENTS

I am deeply grateful to Professor Mark C.K. Yang for being an instrumental part of my education. His understanding and confidence were important in my entry to this graduate program. Throughout my studies, his patience, help and guidance were invaluable to my work. His contribution to this dissertation was crucial.

I am thankful to my family for their constant support: my wife, Hung-Ir, for her patience, help and understanding; my daughter, Shue-Fong, for the joys and the change of pace; and my parents and brothers, who were always there with love and encouragement.

My appreciation extends to Professors John Saw, Andy Rosalsky, and Ken Portier for their great help and friendship.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS.....ii

ABSTRACT.....v

CHAPTER

I INTRODUCTION.....1

1.1 Statement of the Problems.....1
1.2 Literature Review.....3
1.3 Sketch of This Dissertation.....6

II COVERAGE PROBABILITY FOR AR(1) MODEL.....7

2.1 General Asymptotic Results.....7
2.2 Asymptotic Result for Stationary Normal ARMA(p,q) Process.....9
2.3 A Computational Method for Coverage Probability for AR(1) Model.....14

III THE COVERAGE PROBABILITY OF PREDICTION INTERVALS IN AR(1) MODEL.....30

3.1 Computation and Simulation Results.....30
3.2 Applications.....32

IV MODEL BUILDING FOR JITTERY COSINE WAVE.....35

4.1 Introduction.....35
4.2 Description of the Model and Its Stationarity.....38
4.3 Minimum Mean Square Error Predictor.....46
4.4 A Special Case.....49
4.5 Consistent Estimators.....52

V FUTURE RESEARCH: GENERAL MODEL AND SUNSPOT DATA.....60

5.1 General Model.....60
5.2 Modeling Sunspot Data.....62

BIBLIOGRAPHY.....71
BIOGRAPHICAL SKETCH.....74

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

SOME NEW RESULTS ON TWO SIMPLE TIME SERIES MODELS--
PREDICTION COVERAGE FOR AR(1) AND MODEL BUILDING
FOR JITTERY COSINE WAVES

BY

PAN-YU LAI

May, 1985

Chairman: Dr. Mark C.K. Yang
Major Department: Department of Statistics

The coverage probability is defined as the probability of a sequence of random variables falling into a given region simultaneously. Let $\{X_t, t=1,2, \ldots ,k\}$ be a sequence of random variables of size k and $B_t=\{(a_t,b_t), t=1,2, \ldots ,k\}$ be a sequence of intervals in the real line. The exact value of the coverage probability, $Pr\{X_t \epsilon B_t, t=1,2, \ldots ,k\}$, is usually difficult to compute when the X_t's are not independent of each other. In the first part of this dissertation, a computational method to evaluate the coverage probability at any required accuracy when $\{X_t\}$ is a first-order autoregressive normal process or when $\{X_t\}$ is the prediction error sequence of that is derived. A Monte Carlo study to estimate $Pr\{X_t \epsilon B_t, t=1,2, \ldots ,k\}$ was conducted to confirm the computational technique. The results are satisfactorily close to each other.

CHAPTER I
INTRODUCTION

1.1 Statement of the Problems

In this dissertation two slightly related topics in time series are studied. They are related because both of them deal with time series model building. However, they are only slightly related because one deals with model validation in the time domain and the other deals with model building in the frequency domain.

The first problem occurs in time series when one wants to validate the model after he has already picked a time series model and considers it a good approximation to the true time series. A reasonable way to do this is to compute a sequence of prediction values and their corresponding 100 $(1-\alpha)$% confidence intervals and to see whether the proportion of future observations to fall into these intervals agrees with what the confidence intervals predict. Since the confidence intervals are computed separately, the probability $(1-\alpha)$ just applies to each individual forecast but not jointly to the forecasts at the different lead times. It is natural to ask what is the probability that the future time series stays within all the confidence intervals simultaneously. We will call this probability the coverage probability. The coverage probability can also be used to find the first passage time distribution for a stationary process. A computational method to compute the coverage probability for some

special cases of stationary time series is the main theme of the next two chapters.

The second problem occurs when one is dealing with random fluctuations in time series. In analyzing time series data, one probably would like to ask himself the following questions before he starts to fit a model to the data in order to get a good result:

1. Does this set of data show some periodical trend?
2. What kind of trend is it?
3. How shall I deal with the trend if it exists?

The answer to question 1 is yes quite often because time series data more or less show regular fluctuations. These up-and-down movements (or trends) may be strictly periodic or nearly so. They might fluctuate irregularly but always keep the same period in each cycle. Another possible type of irregular fluctuation is that its period varies from cycle to cycle. Hours of sunshine per day is an example of the first type. Examples of the second type are economic time series such as weekly total sales in which the fluctuations reflect or constitute the business cycle or physical time series such as weekly temperatures in which the fluctuations come from the change of the season.

In analyzing data with trend of the first or second type, one always detrends the data by taking the differences between the data or decomposing the series into a sum of a deterministic periodical function such as some trigonometric functions and some random component.

Some examples of time series with the third type of trend are sunspot data and the speech wave-form. This kind of data does possess

some cyclical trend but its frequency varies. Thus, we call it jittery cosine wave. The term jittery comes from linguists and the term cosine wave comes from what it looks like. The methods mentioned in the preceding paragraph are not adequate to analyze this kind of data because these two methods require constant period from time to time. In order to give some idea about how to model this kind of data, we propose a new time series model which allows the period to be different from cycle to cycle.

1.2 Literature Review

Let $\{X_t,\ t \geq 1\}$ be a stationary time series and $B_t = (a_t, b_t)$ $t = 1,2, \ldots$ be a given sequence of intervals in the real line. Also, let the coverage probability P_k be defined as

$$P_k = Pr\{X_t \in B_t \ ;\ t = 1,2,\ldots,k\} \ . \tag{1.2.1}$$

The computation of P_k seems to be an open problem since the unsuccessful attempt by Epstein (1949, 1951). It is related to the general category of the extreme values and the first passage time problems. Surveys of recent development and applications of extreme values can be found in Gumbel (1958), Hann (1976), Galambos (1978), and David (1970). When the X_t's are independently and identically distributed random variables with a known distribution, the values of P_k can be easily calculated, and when X_t's are independent and identically distributed with an unknown distribution, the asymptotic property of P_k evaluated at $a_t = -\infty$ and $b_t = b$ for $t = 1,2, \ldots, k$ can be proved to belong to one of three categories of distributions

for one-sided identical intervals $B_1 = B_2 = \ldots = (-\infty,b)$, or (a,∞). Although a similar asymptotic theory can be derived for correlated X_t (see e.g., Watson, 1954; Berman, 1964, 1971; Leadbetter, 1974; Leadbetter and his colleagues, 1978, 1979), little is known about the nonasymptotic properties of P_k. As one might expect, while the properties of P_k differ very little asymptotically between independent and correlated processes, they may be quite different for small k. Hunter (1977) has found a bound for P_k under the first-order autoregressive model by his modified Bonferroni inequality combined with the results given by Gupta (1963) and Slepian (1962), but it seems difficult to improve his bounds further if the error of their bounds is greater than the required accuracy. The first passage time problem has received even more attention in the literature. A survey of this field can be found in Blake and Lindsey (1973) where 133 references are cited.

Coverage probability can also be used to check the reliability of using prediction intervals for time series model validation. Many authors (e.g., Box and Jenkins, 1976, p. 137-138, 146; Anderson, 1975, p. 122; and Nelson, 1973, p. 153) use prediction intervals to validate their model by checking whether the future observations fall into these intervals. But little is known about their real coverage probabilities. We solved this problem at least for any first-order autoregressive model with positive autoregressive coefficients. Since the prediction intervals are in general two-sided and nonidentical, it is not known whether similar bounds can be easily found by Hunter's method.

Wolfer's yearly sunspot data have been studied by many authors. A variety of models which may be based on different time periods have been suggested since 1927. A summary of these works is given by Woodward and Gray (1978), where nine different autoregressive-moving average models (ARMA) can be found. In their paper, three models suggested by them are compared with and shown superior to a second order autoregressive model which was suggested by Box and Jenkins (1976) and Moran (1954). Although Woodward and Gray claim that their models are better than other ARMA models, they admit that probably the most satisfactory method of modeling the sunspot series would be one which utilized the actual physical mechanisms at work. Without knowing what the actual mechanisms are, we tried to find a model which provides an improved fit to the data. A model which is a product of two jittery cosine waves is, therefore, proposed.

Speech wave-form has been studied by many linguists and acousticians. A vocal jitter, which is a laryngeal phenomenon defined as the cycle-to-cycle variation found within successive periods of a laryngeal vibratory pattern, makes the speech wave-form fail to keep the fixed frequency from cycle-to-cycle. Horii (1975, 1979) analyzed this kind of data by using a peak-picking method which is similar to the one described by Gold (1962) to obtain sample frequencies and used them to estimate the mean, median and standard deviation of voice fundamental frequencies. Since Gold's peak-picking method did not use any modelling techniques and just used a computer program to select the maxima and minima of the wave-form, it would be desirable to incorporate a statistical model in analyzing the fundamental frequencies.

1.3 Sketch of This Dissertation

In Chapter II, a computational method is developed to compute P_k of a first-order autoregressive process to any required accuracy. Of course, the greater the accuracy is required, the more computation is necessary. However, the computation time is approximately linear in k for any fixed accuracy. This method then will be used to compute the coverage probability of the prediction errors of a first-order autoregressive process. One hundred and twenty coverage probabilities of the various numbers of prediction errors of first-order autoregressive process with 12 different autoregressive coefficients are computed by using our method in Chapter III as an example. Since the true coverage probabilities are unknown, we also conducted a Monte Carlo study in Chapter III to compare with the computational results. (Some applications, such as first passage time distribution, are also discussed in this chapter.)

A new model which contains a random frequency shift in a cosine function for a stationary time series is introduced and studied in Chapter IV. A set of consistent estimators are also given in Chapter IV. In Chapter V, we state some future research in this area which includes the extension of this model to a more general case and the construction of a more complicated model to fit the Wolfer's yearly sunspot data. Our model is shown better than those models suggested by Woodward and Gray (1978) because it can fit the sample autocorrelations of sunspot data much better.

CHAPTER II
COVERAGE PROBABILITY FOR AR(1) MODEL

2.1 General Asymptotic Results

Let $M_n = \max_{1 \le i \le n} \{X_i\}$. Then the coverage probability P_n defined in (1.2.1) evaluated at $B_t = (-\infty, b)$ for a real number b and for t = 1,2, . . . ,n, referred to as $P_{n,b}$ from now on, is the same as $Pr(M_n < b)$ which is an extreme value distribution function. The classical extreme value theory which deals with the asymptotic distribution for the maximum of independent and identically distributed (i.i.d.) random variables has been studied for at least half a century. However, some special cases may reach further back to mathematical antiquities. For example, as early as in 1709, Nicolous Bernoulli considered the following actuarial problem: Assume n men of equal age will die within t years; what is the mean duration of life of the last survivor?

The classical results about extreme value distribution can be found in many articles. For example, in Galambos' book (1978), it is stated as follows:

If $X_1, X_2, \ldots$ are i.i.d. random variables, and if for some real sequences $a_n > 0$, b_n, the normalized maximum $a_n(M_n - b_n)$ has a non-degenerated limiting distribution function G(x), then G(x) has one of the following forms:

TYPE I: $G(x) = \exp(-e^{-x})$ $\quad -\infty < x < \infty$,

TYPE II: $G(x) = 0$ $\quad x \leq 0$

$= \exp(-x^{-\alpha})$ (for some $\alpha > 0$) $\quad x > 0$,

TYPE III: $G(x) = \exp(-(-x)^{\alpha})$ (for some $\alpha > 0$) $\quad x \leq 0$

$= 1$ $\quad x > 0$.

It is known that if $X_1, X_2, \ldots$ are independent and identically distributed normal random variables with zero mean and unit variance, then the asymptotic distribution of M_n is of type I, i.e.,

$$\Pr(a_n(M_n - b_n) \leq x) \to \exp(-e^{-x}) \text{ , as } n \to \infty \text{ ,}$$

where

$$a_n = (2\log(n))^{1/2} \tag{2.1.1}$$

$$b_n = (2\log n)^{1/2} - 0.5(2\log n)^{-1/2}\{\log\log n + \log(4\pi)\} \text{ .} \tag{2.1.2}$$

Berman (1964) extended this result to a stationary normal sequence and showed that under certain conditions, the extreme value of a stationary normal sequence has the same limiting distributions as in the case of i.i.d. random variables. His results are given in the following theorem.

Theorem 2.1.1. Let $\{X_n, n=1,2,\ldots\}$ be a stationary normal sequence with

$$EX_n = 0,\ EX_n^2 = 1 \tag{2.1.3}$$

for all n and with an autocovariance sequence $\{r_k,\ k=1,2,\ldots\}$ defined by

$$r_k = EX_1X_{1+k} \qquad k=1,2,\ldots. \tag{2.1.4}$$

If $\{r_k\}$ satisfies

$$\lim_{n\to\infty} r_n \log n = 0 \tag{2.1.5}$$

or

$$\sum_{n=1}^{\infty} r_n^2 < \infty \tag{2.1.6}$$

then $M_n = \max\{X_1, X_2, \ldots, X_n\}$ satisfies

$$\Pr(a_n(M_n - b_n) \leq x) \to \exp(-e^{-x}), \quad \text{as } n \to \infty,$$

where a_n and b_n are defined as in (2.1.1) and (2.1.2).
This theorem will be used later to find the asymptotic distribution for the maximum of a stationary normal ARMA(p,q) process. The definition of an ARMA(p,q) process and all the details are given in the next section.

2.2 Asymptotic Result for Stationary Normal ARMA(p,q) Process

A mixed autoregressive-moving average (ARMA) process $\{X_t\}$ is defined as

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \qquad (2.2.1)$$

where the ϕ's and θ's are real constants; p,q are two nonnegative integers and $\{a_t\}$ are i.i.d. normal random variables. Moreover, the future noise a_{t_0} is independent of the past X_t for $t < t_0$ for all t_0. In other notation, let B be the backward shift operator which means $BX_t = X_{t-1}$ for all t. Then

$$(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)X_t = (1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)a_t$$

or

$$\phi(B)X_t = \theta(B)a_t$$

where $\phi(B)$ and $\theta(B)$ are polynomials of degree p and q in B.

An even simpler notation for process (2.2.1) is ARMA(p,q). If the equation $\phi(B) = 0$ has all its roots lying outside the unit circle, then the process defined in (2.2.1) is a stationary process.

Now, let the X_t's satisfy condition (2.1.3) and r_n be defined in (2.1.4). Then the autocovariance function satisfies

$$r_k = \phi_1 r_{k-1} + \ldots + \phi_p r_{k-p} + r_{Xa}(k) - \theta_1 r_{Xa}(k-1) - \ldots - \theta_q r_{Xa}(k-q) \qquad (2.2.2)$$

where $r_{Xa}(k)$ is the cross covariance function between X and a and is defined by

$$r_{Xa}(k) = E(X_{t-k} a_t) .$$

By the independence of the future $\{a_t\}$ and the past $\{X_t\}$, equation

(2.2.2) implies the difference equation

$$r_k = \phi_1 r_{k-1} + \phi_2 r_{k-2} + \ldots + \phi_p r_{k-p} \qquad k \geq q+1 \qquad (2.2.3)$$

or

$$\phi(B) r_k = 0 \qquad k \geq q+1$$

with initial conditions $\{r_k,\ k=1,2,\ldots,q\}$ which depend on the q moving average parameters $(\theta_1,\theta_2,\ldots,\theta_q)$ and the p autoregressive parameters $(\phi_1,\phi_2,\ldots,\phi_p)$. The general solution of (2.2.3) is

$$r_n = G_1^n P_{m_1-1}(n) + \ldots + G_s^n P_{m_s-1}(n)$$

where $\phi(B) = 0$ has roots

G_1 with multiplicity m_1,

$\vdots$

G_s with multiplicity m_s,

$m_1 + m_2 + \ldots + m_s = P$ and

$P_i(n)$ is an arbitrary polynomial of order i in n.

It can be shown that for a stationary process, G_i's satisfy

$$|G_i| < 1 \qquad i=1,2,\ldots,p\ .$$

Therefore, if $\{X_n,\ n=1,2,\ldots\}$ is a stationary normal ARMA(p,q) process satisfying condition (2.1.3), then it can be proved that

$$\sum_{k=1}^{\infty} r_k^2 = \sum_{k=1}^{\infty} (G_1^n P_{m_1-1}(n) + \ldots + G_s^n P_{m_s-1}(n))^2$$

$$< \infty .$$

Thus, the condition (2.1.6) of Theorem 2.1.1 is satisfied. Hence, we can conclude that

$$\Pr(a_n(M_n - b_n) \leq x) \to \exp(-e^{-x}) \quad \text{as } n \to \infty \tag{2.2.4}$$

where a_n and b_n are defined as in (2.1.1) and (2.1.2). The coverage probability P_n of a stationary normal ARMA(p,q) process can be approximated through (2.2.4) when $B_t = (-\infty, b)$ and n is large. But can (2.2.4) be used for small n? To answer this question, a Monte Carlo study for one of the simplest models described in (2.2.1), by letting $p = 1$ and $q = 0$, known to as AR(1) process, is conducted. The study is done by generating 10,000 simulations for each different combination of ϕ_1 and n values. In each simulation, a set of n normally distributed observations, $X_1, X_2, \ldots, X_n$, which satisfies condition (2.1.3) is randomly generated through AR(1) model with autoregressive parameter ϕ_1. Then the coverage probability $P_{n,1.645}$ is estimated by C/10,000 where C is the number of simulations of which the whole set of observation values falls into the 95% one-sided confidence interval; that is, $X_t < 1.645$, $t=1,2,\ldots,n$. The result is shown in Table I. From Table I, we can see that when n is small, the coverage probabilities of the AR(1) process when the autoregressive

parameter (ϕ_1) is not close to zero are quite different from those in the i.i.d. case. For example, when k = 10

$$P_{10,1.645} = 0.5988 \qquad \text{at } \phi_1 = 0 \qquad \text{and}$$

$$P_{10,1.645} = 0.7719 \qquad \text{at } \phi_1 = 0.9 \ .$$

We can claim that their difference is statistically significant because two standard deviations of the simulation error based on 10,000 independent Bernoulli trials is less than 0.01.

Table I

$P_{n,1.645}$ from 10,000 Simulations

ϕ_1	n = 4	5	8	10
0	0.8145	0.7738	0.6634	0.5988
0.02	0.8120	0.7716	0.6649	0.6009
0.1	0.8186	0.7751	0.6646	0.6144
0.2	0.8174	0.7769	0.6707	0.6139
0.5	0.8271	0.7961	0.7080	0.6671
0.9	0.8158	0.8017	0.7876	0.7719

2.3 A Computational Method for Coverage Probability for AR(1) Model

Since the result at the end of the previous section shows that when n is not large, (2.2.4) cannot be applied to compute $P_{n,b}$, methods other than (2.2.4) need to be explored when n is not large. Hunter (1977) has found a bound for $P_{n,b}$ for an AR(1) normal process by his modified Bonferroni inequality combined with the results by Gupta (1963) and Slepian (1962), but it seems difficult to improve his bounds further if the width of their bounds is greater than the required accuracy. Moreover, the prediction intervals are in general two-sided and nonidentical. It is not clear whether similar bounds can be easily found by Hunter's method. In the rest of this section, we derive a computational method that can compute P_k for any required accuracy for an AR(1) normal process or a conditional AR(1) normal process (which will be defined later) for both one-sided and two-sided nonidentical intervals. Of course, the greater the accuracy is required, the more computation is necessary.

We first define the conditional AR(1) normal process as follows: Let X_t be a Markov process satisfying $X_0 = 0$ and

$$X_t = \phi X_{t-1} + \varepsilon_t , \qquad t=1,2,... \tag{2.3.1}$$

where $|\phi| < 1$ and $\{\varepsilon_t\}$ is a sequence of i.i.d. normal random variables with zero mean and unit variance. Also let $B_t = (a_t, b_t)$, $t=1,2,...,$ be a given sequence of intervals in the real line. In this section,

$$P_k = Pr(X_t \in B_t \ ; \ t=1,2,...,k)$$

is evaluated for small to moderate k when ϕ is a rational number of the form

$$\phi = s/r$$

where $s,r \neq 0$ are integers and s/r is irreducible.

Let a new discrete process $X_t(n)$ and $\varepsilon_t(n)$ be defined as

$$X_t(n) = \phi X_{t-1}(n) + \varepsilon_t(n) \qquad (2.3.2)$$

where $X_0(n) \equiv 0$, and $\{\varepsilon_t(n)\}$ is an i.i.d. discrete process with

$$Pr\{\varepsilon_t(n) = jr^{-n}\} = \Phi[(j+1/2)r^{-n}] - \Phi[(j-1/2)r^{-n}], \quad j=0,\pm1,\ldots,$$

where $\Phi(x)$ is the standard normal distribution function.

Equation (2.3.2) is the result of approximating ε_t by a discrete random variable $\varepsilon_t(n)$ which condenses all the probability in $[(j-1/2)r^{-n}, (j+1/2)r^{-n}]$ to jr^{-n}. To use this approximation, one needs to control the error

$$e_k = Pr\{X_t \varepsilon B_t;\ t=1,2,\ldots,k\} - Pr\{X_t(n) \varepsilon B_t;\ t=1,2,\ldots,k\}\ , \qquad (2.3.3)$$

and to find a feasible algorithm to compute $Pr\{X_t(n) \varepsilon B_t,$ $t=1,2,\ldots,k\}$. The second problem is dealt with first. The apparent difficulty in finding $Pr\{X_t(n) \varepsilon B_t;\ t=1,2,\ldots,k\}$ lies in the fact that $X_t(n)$ takes more and more values for B_t of the same length. For example, $X_1(n)$ takes only values of the form $jr^{-n} \varepsilon B_1$ for integer j,

while $X_2(n)$, according to (2.3.2) takes values of the form $jr^{-(n+1)}\varepsilon B_2$ for integer j. Consequently, $X_k(n)$ takes values of the form $jr^{-(n+k-1)}\varepsilon B_k$ for integer j. Even for moderate k (e.g., k=10), and small n, r (e.g., n=4, r=5), the storage for the values of $X_k(n)$ in [0,1] is of the order $5^{13} = 1.22 \times 10^9$ which cannot be handled by most computers. Because of the large number of values for $X_k(n)$, the time required to compute them is even worse. Let $P_i(n,j)$ be the probability that $X_i(n) = jr^{-(n+i-1)}$. Then a recursive formula from $P_{i-1}(n,j)$ to $P_i(n,j)$ can be written as

$$P_i(n,j) = \sum_{j'} Pr[\varepsilon_i(n)=jr^{-(n+i-1)}-\phi j' r^{-(n+i-2)}]p_{i-1}(n,j'), \quad (2.3.4)$$

where the range of j' is over all the integer j' such that $j'r^{-(n+i-2)}\varepsilon B_{i-1}$ and $Pr\{\varepsilon_i(n) = jr^{-(n+i-1)}-\phi j' r^{-(n+i-2)}\} = 0$ if $jr^{-(n+i-1)}-\phi j' e^{-(n+i-2)}$ is not of the form ℓr^{-n} for some integer ℓ. The total additions and multiplications for computing P_k by (2.3.4) behave like a k-fold multiple integral, i.e., with total computational operations $O(r^{2k})$. This is obviously unmanageable even for small k. The following theorem will enable one to reduce the storage space as well as the computing time to O(k) for fixed n and r.

Theorem 2.3.1. Let $\{X_t\}$ denote the time series defined in (2.3.1). Then

$$|Pr\{X_t\varepsilon B_{1t},\ t=1,2,\ldots,k\} - Pr\{X_t\varepsilon B_{2t},\ t=1,2,\ldots,k\}| \le Ar^{-(2n+k-1)}, \quad (2.3.5)$$

where

$$A = \frac{s}{\pi r} \cdot (1-(\tfrac{s}{r})^2)/(1-(\tfrac{s}{r})^{2k})+(2\pi e)^{-1/2}\,(1+\frac{s^2}{r^2}(1-(\tfrac{s}{r})^{2(k-1)})/(1-(\tfrac{s}{r})^{2k})),$$

$$B_{1t} = B_{2t} = B_t \quad \text{for } t=1,2,\ldots,k-1, \text{ and}$$

$$B_{1k} = [j/r^n - 1/(2r^{n+k-1}), j/r^n + 1/(2r^{n+k-1})] ,$$

$$B_{2k} = [j/r^n - (m+\tfrac{1}{2})/r^{n+k-1}, j/r^n - (m-\tfrac{1}{2})/r^{n+k-1}] ,$$

for any $j=0,\pm1,\pm2,\ldots,$ and any $0 < m < r^{k-1}$.

Proof: Let the density function of $\underset{\sim}{X} = (X_1,X_2,\ldots,X_k)'$ be denoted by

$$f(\underset{\sim}{X}) = (2\pi)^{-k/2}|\underset{\sim}{\Sigma}|^{-1/2} \exp(-\underset{\sim}{X}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{X}/2) ,$$

where $\underset{\sim}{\Sigma}$ is the variance-covariance matrix of $\underset{\sim}{X}$, i.e.,

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{k-1} \\ \phi & 1+\phi^2 & \phi(1+\phi^2) & \cdots & \phi^{k-2}(1+\phi^2) \\ \vdots & & & & \vdots \\ \phi^{k-1} & & \cdots & & 1+\phi^2+\ldots+\phi^{2(k-1)} \end{bmatrix} .$$

Then it is known by Theorems 8.3.6-7 of Graybill (1969), that the last column and row of Σ^{-1} can be expressed by

$$\sigma^{ik} = \sigma^{ki} = \begin{cases} 1 & \text{if } i=k \\ -s/r & \text{if } i=k-1 \\ 0 & \text{otherwise,} \end{cases} \tag{2.3.6}$$

where σ^{ij} denotes the i,jth entry of $\underset{\sim}{\Sigma}^{-1}$. Therefore,

$$\begin{aligned} &|\Pr\{X_t \epsilon B_t,\ t=1,2,\ldots,k\} - \Pr\{X_t \epsilon B_{2t},\ t=1,2,\ldots,k\}| \\ &\quad = |\int_{B_1}\int_{B_2}\cdots\int_{B_{k-1}}\int_{B_{1k}} f(\underset{\sim}{x})d\underset{\sim}{x} - \int_{B_1}\int_{B_2}\cdots\int_{B_{2k}} f(\underset{\sim}{x})d\underset{\sim}{x}| \\ &\quad = |\int_{B_{1k}} g(x_k)dx_k - \int_{B_{2k}} g(x_k)dx_k| \\ &\quad \leq \max_{u \epsilon B_{1k},\ v \epsilon B_{2k}} |g(u)-g(v)|r^{-(n+k-1)}\ , \end{aligned} \tag{2.3.7}$$

where $g(x_k) = \int_{B_1}\cdots\int_{B_{k-1}} f(\underset{\sim}{x})dx_{k-1}dx_{k-2}\cdots dx_1$. The right hand side (2.3.7) can be simplified by noting that

$$\max|g(u)-g(v)| \leq \max_{w_0 \epsilon I_0} |g'(w_0)|r^{-n}\ , \quad \text{where}$$

$I_0 = [jr^{-n}-(1/2)r^{-(n+k-1)}, jr^{-n}+(m+1/2)r^{-(n+k-1)}]$, and

$$|g'(w_0)| = |\frac{\partial}{\partial x_k}\int_{B_1}\cdots\int_{B_{k-1}} f(\underset{\sim}{x})dx_1dx_2\ldots dx_{k-1}|$$

$$= \left| \int_{B_1} \cdots \int_{B_{k-1}} \frac{\partial}{\partial x_k} f(\underset{\sim}{x}) dx_1 \cdots dx_{k-1} \right|$$

$$= \left| \int_{B_1} \cdots \int_{B_{k-1}} (-\tfrac{1}{2} f(\underset{\sim}{x}) (\sum_{i=1}^{k-1} x_i (\sigma^{ik} + \sigma^{ki}) + 2x_k \sigma^{kk}) dx_1 \cdots dx_{k-1} \right| . \tag{2.3.8}$$

Substituting (2.3.6) to (2.3.8), one has

$$|g'(w_0)| \leq (s/r) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |x_{k-1}| f(\underset{\sim}{x}) dx_1 \cdots dx_{k-1} + |x_k| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\underset{\sim}{x}) dx_1 \cdots dx_{k-1}$$

$$= (s/r) \int_{-\infty}^{\infty} f(x_{k-1}, x_k) |x_{k-1}| dx_{k-1} + |x_k| f(x_k)$$

and $\int_{-\infty}^{\infty} f(X_{k-1}, X_k) |X_{k-1}| dX_{k-1}$

$$= \int_{-\infty}^{\infty} f(x_k) f(x_{k-1} | X_k = x_k) \cdot |X_{k-1} - E(X_{k-1} | X_k) + E(X_{k-1} | X_k)| dX_{k-1}$$

$$\leq f(x_k) \int_{-\infty}^{\infty} f(X_{k-1} | X_k = x_k) |X_{k-1} - E(X_{k-1} | X_k)| dX_{k-1} +$$

$$f(X_k) \int_{-\infty}^{\infty} f(X_{k-1} | X_k) |E(X_{k-1} | X_k)| dX_{k-1}$$

$$= f(X_k) \cdot \frac{2}{\pi} + f(X_k) |X_k| \cdot \operatorname{cov}(X_k, X_{k-1}) / \operatorname{var}(X_k)$$

$$\leq \frac{2}{\pi} / (2\pi \cdot \operatorname{var}(X_k))^{1/2} + f(X_k) |X_k| \frac{s}{r} \cdot (1 - (\tfrac{s}{r})^{2(k-1)}) / (1 - (\tfrac{s}{r})^{2k}) ,$$

$$\max_{-\infty < X_k < \infty} f(X_k) |X_k| = (2\pi e)^{-1/2} ,$$

we have

$$|g'(w_0)| < \frac{s}{r}[\frac{1}{\pi} \cdot \quad (1-(\frac{s}{r})^2)/(1-(\frac{s}{r})^{2k}) \quad +$$

$$\frac{s}{r} \cdot (2\pi e)^{-1/2} \cdot (1-(\frac{s}{r})^{2(k-1)})/(1-(\frac{s}{r})^{2k})] + (2\pi e)^{1/2} \ .$$

This completes the proof of Theorem 2.3.1.

To see how Theorem 2.3.1 can be used to reduce storage and computation, we let $B_1 = B_2 \ . \ . \ . = B_k = [0,1]$ to simplify the illustration. Without Theorem 2.3.1, one needs all the probability values for $X_k(n)$ at points

$$\{0, r^{-(n+k-1)}, 2r^{-(n+k-1)}, \ldots, jr^{-(n+k-1)}, \ldots, 1\} \tag{2.3.9}$$

to approximate P_k by (2.3.4). But with Theorem 2.3.1, one needs only to compute $X_k(n)$ at

$$\{0, r^{-n}, 2r^{-n}, \ldots, jr^{-n}, \ldots, 1\} \tag{2.3.10}$$

and approximate all the rest of the probabilities at points in (2.3.9) by the probabilities at points in (2.3.10). To compute the probabilities of points in (2.3.10) by (2.3.4), note that in (2.3.4) $\varepsilon_k(n)$ takes only values of the form ir^{-n} with integer i. Thus, for the point jr^{-n} in (2.3.10), one has, by (2.3.4),

$$ir^{-n} = jr^{-n} - (s/r)j'r^{-(n+k-2)} \ , \tag{2.3.11}$$

or

$$j' = (j-i)r^{k-1}/s \ .$$

Because s and r are relative primes, (j-i)/s has to be an integer. In other words, $j'r^{-(k-1)}$ has to be an integer. Thus, for all $j'r^{-(n+k-2)} \varepsilon B_{k-1}$, one only needs to consider the values $\{0, r^{-n+1}, 2r^{-n+1}, \ldots, jr^{-n+1}, \ldots, 1\} \varepsilon B_{k-1}$. It can be shown by the same argument that in order to compute all the probabilities at the points in (2.3.10), one needs only to evaluate the probabilities of all the subsequent $X_{k-2}(n), X_{k-3}(n), \ldots, X_2(n), X_1(n)$ at $\{0, r^{-n+1}, 2r^{-n+1}, \ldots, jr^{-n+1}, \ldots, 1\}$. Hence, the total number of operations is of order $O(kr^{2n-1})$ and the storage space is $O(r^n)$. For fixed n, the computational time is linear in k.

Next, the error defined by (2.3.3) will be evaluated.

Theorem 2.3.2. Let X_t and $X_t(n)$ be defined by (2.3.1) and (2.3.2), respectively. Then, the error e_k defined by (2.3.3) satisfies

$$|e_k| \leq k(k+1)e_0/2$$

where $e_0 = (2\pi)^{-\frac{1}{2}}(r^n)^{-1} \cong 0.4r^{-n}$.

Before proving this theorem, we need to prove a lemma. In this lemma, [x] will denote the closest integer to x, e.g., [1.4] ≡ 1, [1.6] ≡ 2, [1.5] ≡ 1.

Lemma 2.3.3. Let $\{p_j\}$ be a sequence of nonnegative numbers such that $\sum_{j=-\infty}^{\infty} p_j \leq 1$. Let $\max_{-\infty<j<\infty} |p_j| \leq e_0$, and for a given integer k,

$$S_k = \sum_{j_1\varepsilon[a_1,b_1]} \sum_{j_2\varepsilon[a_2,b_2]} \cdots \sum_{j_k\varepsilon[a_k,b_k]} p_{j_1} p_{j_2} \cdots p_{j_k}$$

$$S_k' = \sum_{j_1\varepsilon[a_1',b_1']} \sum_{j_2\varepsilon[a_2',b_2']} \cdots \sum_{j_k\varepsilon[a_k',b_k']} p_{j_1} p_{j_2} \cdots p_{j_k},$$

where a_i, b_i, a_i', b_i' may depend on the previous indices $j_{i-1}, j_{i-2}, \ldots, j_1$. If $|a_i - a_i'| \le 1$, $|b_i - b_i'| \le 1$, and $a_i \ge a_i'$ if $b_i \ge b_i'$, for all $i=1,2,\ldots,k$, then $|e_k| \equiv |S_k - S_k'| \le k e_0$.

The proof is obvious by induction and by the general formula

$$\sum_{j_1=a_1}^{b_1} \sum \cdots \sum_{j_k=a_k}^{b_k} p_{j_1} \cdots p_{j_k} - \sum_{j_1=a_1'}^{b_1'} \sum \cdots \sum_{j_k=a_k'}^{b_k'} p_{j_1} \cdots p_{j_k}$$

$$= \sum_{j_1=a_1}^{b_1} \Big[\sum_{a_2}^{b_2} \cdots \sum_{a_k}^{b_k} p_{j_2} \cdots p_{j_k} - \sum_{a_2'}^{b_2'} \cdots \sum_{a_k'}^{b_k'} p_{j_2} \cdots p_{j_k} \Big] p_{j_1} + X \sum_{a_2'}^{b_2'} \cdots \sum_{a_k'}^{b_k'} p_{j_2} \cdots p_{j_k},$$

where $|X| \le \max(p_{b_1}, p_{a_1}, p_{b_1'}, p_{a_1'}, |p_{b_1} - p_{a_1}|, |p_{b_1'} - p_{a_1'}|) \le e_0$.

Proof of Theorem 2.3.2.

In this proof, all the j_i's are r^{-n} integer; i.e., they are of the form jr^{-n} for an integer j. Let $[x]$ denote the number jr^{-n} closest to x and $\Delta = r^{-n}/2$. By the transformation $X_t = \phi X_{t-1} + \varepsilon_t$, $t=1,2,\ldots,k$, we have

$$\Pr\{X_t \varepsilon B_t,\ t=1,2,\ldots,k\}$$

$$= \int_{\alpha_1}^{\beta_1} \int_{\alpha_2}^{\beta_2} \cdots \int_{\alpha_k}^{\beta_k} d\phi(\varepsilon_k) d\phi(\varepsilon_{k-1}) \cdots d\phi(\varepsilon_1),$$

where $\alpha_1 = a_1$, $\beta_1 = b_1$, $\alpha_2 = a_2 - \phi\varepsilon_1$, $\beta_2 = b_2 - \phi\varepsilon_1$, and in general

$$\alpha_\ell = a_\ell - \phi\varepsilon_{\ell-1} - \ldots - \phi^{\ell-1}\varepsilon_1 ,$$

$$\beta_\ell = b_\ell - \phi\varepsilon_{\ell-1} - \ldots - \phi^{\ell-1}\varepsilon_1 , \quad \ell=1,2,\ldots,k. \tag{2.3.12}$$

Define

$$I_i \equiv I_i(\varepsilon_1,\ldots,\varepsilon_{k-i}) \equiv \int_{\alpha_{k-i+1}}^{\beta_{k-i+1}} \cdots \int_{\alpha_k}^{\beta_k} d\phi(\varepsilon_k)\ldots d\phi(\varepsilon_{k-i+1}) ,$$

$$S_i \equiv S_i(\varepsilon_1,\ldots,\varepsilon_{k-i}) \equiv \sum_{j_{k-i+1}=A_{k-i+1}}^{B_{k-i+1}} \cdots \sum_{j_k=A_k}^{B_k} Pr(\varepsilon_k(n)=j_k)$$

$$\cdot\, Pr(\varepsilon_{k-1}(n)=j_{k-1}) \ldots Pr(\varepsilon_{k-i+1}(n)=j_{k-i+1}),$$

where

$$A_\ell = [a_\ell - \phi\varepsilon'_{\ell-1} - \phi^2\varepsilon'_{\ell-2} - \ldots - \phi^{\ell-1}\varepsilon'_1]$$

$$B_\ell = [b_\ell - \phi\varepsilon'_{\ell-1} - \phi^2\varepsilon'_{\ell-2} - \ldots - \phi^{\ell-1}\varepsilon'_1]$$

with

$$\varepsilon'_m = \begin{cases} \varepsilon_m & \text{if } m \le k-i \\ j_m & \text{if } m > k-i . \end{cases}$$

Then

$|\alpha_k - A_k| \le \Delta/2$, $|\beta_k - B_k| \le \Delta/2$, and

$$I_1 - S_1 = \int_{\alpha_k}^{\beta_k} d\Phi(\varepsilon_k) - \sum_{j_k=A_k}^{B_k} Pr(\varepsilon_k(n) = j_k)$$

$$= \int_{\alpha_k}^{\beta_k} d\Phi(\varepsilon_k) - \int_{A_k}^{B_k} d\Phi(\varepsilon_k) \leq (\Delta/2 + \Delta/2) \cdot \frac{1}{(2\pi)^{1/2}} \cdot r^{-n} = e_0 \cdot$$

Let $||I_i - S_i|| = \sup_{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{k-1}} |I_i - S_i|$ and

$$S^*_{i-1} = \sum_{j_{k-i+2}=A'_{k-i+2}}^{B'_{k-i+2}} \sum_{j_{k-i+3}=A'_{k-i+3}}^{B'_{k-i+3}} \cdots$$

$$\sum_{j_k=A_k}^{B'_k} Pr(\varepsilon_k(n)=j_k) \ldots Pr(\varepsilon_{k-i+2}(n)=j_{k-i+2}) ,$$

where

$$A'_{k-i+2} = [a_{k-i+2} - \phi[\varepsilon_{k-i+1}] - \phi^2 \varepsilon_{k-i} - \cdots - \phi^{k-i+1}\varepsilon_1]$$

$$A'_{k-i+3} = [a_{k-i+3} - \phi j_{k-i+2} - \phi^2[\varepsilon_{k-i+1}] - \phi^3 \varepsilon_{k-i} - \cdots - \phi^{k-i+2}\varepsilon_1]$$

$$\vdots$$

$$A'_k = [a_k - \phi j_{k-1} - \cdots - \phi^{i-2} j_{k-i+2} - \phi^{i-1}[\varepsilon_{k-i+1}] - \phi^i \varepsilon_{k-i} - \cdots - \phi^{k-1}\varepsilon_1].$$

Note that A'_ℓ and A_ℓ have a difference at most $1 (r^{-n}$ integer) and $A_\ell \geq A'_\ell$ if $B_\ell \geq B'_\ell$ and

$$\int_{a-\Delta}^{b+\Delta} h([\varepsilon],\ldots)\, d\Phi(\varepsilon) = \sum_{i=a}^{b} h(i,\ldots) Pr(\varepsilon(n) = i),$$

for any function h and r^{-n} integers a and b. Thus, for the general term

$$I_i - S_i = \int_{\alpha_{k-i+1}}^{\beta_{k-i+1}} (I_{i-1} - S_{i-1}) + (S_{i-1} - S^*_{i-1})\, d\Phi(\varepsilon_{k-i+1})$$

$$+ \int_{\alpha_{k-i+1}}^{\beta_{k-1+1}} S^*_{i-1} d\Phi(\varepsilon_{k-i+1}) - S_i \text{, and} \qquad (2.3.13)$$

its second right term satisfies Lemma 2.3.3 and produces $||S_{i-1} - S^*_{i-1}|| \leq (i-1)e_0$. The last term of (2.3.13) can be expressed as

$$\delta = \int_{\alpha_{k-i+1}}^{\beta_{k-i+1}} S^*_{i-1} d\Phi(\varepsilon_{k-i+1}) - \int_{A_{k-i+1}}^{B_{k-i+1}} S^*_{i-1} d\Phi(\varepsilon_{k-i+1})$$

and by checking the integration boundaries it can be shown that $|\delta| < e_0$. Thus, (2.3.13) becomes

$$||I_i - S_i|| \leq ||I_{i-1} - S_{i-1}|| + i \cdot e_0 \,,$$

or

$$|I_k - S_k| < k(k+1)e_0/2 \,.$$

The theorem is proven.

An error bound linear in k can be obtained if B_t's are one-sided and $\phi \geq 0$.

Theorem 2.3.4. Let B_t in Theorem 2.3.2 be $(-\infty,b_t)$ for t=1,2,...,k. Then for $\phi \geq 0$, $|e_k| \leq (1.5k-1)e_0$.

First, a lemma will be proven. In this lemma [x] again denotes the closest integer to x.

Lemma 2.3.5. Let F and g be two monotonic functions from $(-\infty,\infty)$ to [0,1]. Let F also be continuous and satisfy $\max_x |F(x+1) - F(x)| \leq e_0$ for some constant e_0.

Then

$$\left|\int_{-\infty}^{a} g(t)dF(t) - \sum_{j=-\infty}^{[a-1/2]} g(j)\{F(j+\tfrac{1}{2}) - F(j-\tfrac{1}{2})\}\right| \leq 1.5\, e_0 . \quad (2.3.14)$$

Proof. First, we write the integral as

$$\int_{-\infty}^{a} g(t)dF(t) = \sum_{j=-\infty}^{[a-1/2]} \int_{j-\frac{1}{2}}^{j+\frac{1}{2}} g(t)dF(t) + \int_{[a-1/2]+\frac{1}{2}}^{a} g(t)dF(t). \quad (2.3.15)$$

Obviously,

$$\int_{j-\frac{1}{2}}^{j+\frac{1}{2}} g(t)dF(t) = m_j\{F(j+\tfrac{1}{2}) - F(j-\tfrac{1}{2})\}$$

for some $m_j \in [g(j-\tfrac{1}{2}), g(j+\tfrac{1}{2})]$. By the monotonicity of g,

$$\left|\int_{j-\frac{1}{2}}^{j+\frac{1}{2}} g(t)dF(t) - g(j)\{F(j+\tfrac{1}{2}) - F(j-\tfrac{1}{2})\}\right| \leq |g(j+\tfrac{1}{2}) - g(j-\tfrac{1}{2})|e_0 . \quad (2.3.16)$$

Combining (2.3.15), (2.3.16), and the fact $0 \le g \le 1$, we have (2.3.14).

Proof of Theorem 2.3.4. As in the proof of Theorem 2.3.2, all the j_i's are of the form jr^{-n} for integer j, and [x] denotes the number jr^{-n} closest to x. Moreover, by the transformation $X_t = \phi X_{t-1} + \varepsilon_t$, we have

$$\Pr\{X_t \varepsilon B_t,\ t=1,2,\ldots,k\}$$

$$= \int_{-\infty}^{b_1} \int_{-\infty}^{b_2-\phi\varepsilon_1} \cdots \int_{-\infty}^{b_k-\phi\varepsilon_{k-1}\cdots-\phi^{k-1}\varepsilon_1} d\phi(\varepsilon_k) \cdots d\phi(\varepsilon_1) .$$

Approximating the last integral by

$$\sum_{j_k=-\infty}^{[b_k-\phi\varepsilon_{k-1}\cdots-\phi^{k-1}\varepsilon_1]} \Pr(\varepsilon_k(n)=j_k) , \qquad (2.3.17)$$

one can see that the difference of (2.3.17) and the last integral is less than $0.5\, e_0$ in absolute value. Let the sum in (2.3.17) be $g(\varepsilon_{k-1})$. Since $\phi \ge 0$, $g(\varepsilon_{k-1})$ is a monotonically decreasing function satisfying the requirement of g(x) in Lemma 2.3.5. Thus, if we approximate

$$\int_{-\infty}^{b_{k-1}-\phi\varepsilon_{k-2}\cdots-\phi^{k-2}\varepsilon_1} \sum_{j_k=-\infty}^{[b_k-\phi\varepsilon_{k-1}-\cdots-\phi^{k-1}\varepsilon_1]} \Pr(\varepsilon_k(n)=j_k) d\phi(\varepsilon_{k-1})$$

by

$$\sum_{j_{k-1}=-\infty}^{\Sigma}[b_{k-1}-\phi\varepsilon_{k-2}-\ldots-\phi^{k-2}\varepsilon_1]\sum_{j_k=-\infty}^{\Sigma}[b_k-\phi j_{k-1}-\ldots-\phi^{k-1}\varepsilon_1]\Pr(\varepsilon_k(n)=j_k)\Pr(\varepsilon_{k-1}(n)=j_{k-1}),$$

(2.3.18)

one can see from Lemma 2.3.5 that the absolute error is less than 1.5 e_0. Similarly, (2.3.18) is a monotonic function of ε_{k-2} and the third integral can again be approximated by a summation with an absolute error less than 1.5 e_0. Thus, by continuing this process, one has the total error less than $(1.5k-1)e_0$ in absolute value. The theorem is proven.

When the three theorems are combined, the absolute error between P_k and (2.3.4) is bounded by $k(k+1)e_0/2 + Ar^{-(2n+k-1)}(b_k-a_k)r^{(n+k-1)}$ $= k(k+1)e_0/2 + A(b_k-a_k)r^{-n}$ for two-sided bounds B_t. Although this error is quadratic in k, it is much better than the usual exponential error propagation in multiple integrals. As shall be shown in the next chapter, the actual error can be much smaller than this upper bound. For one-sided B_t, the error bound becomes $(1.5k-1)e_0 + A\bar{B}_k r^{-n}$, where $\bar{B}_k = (b_k-a_k^*)$ and a_k^* is a lower truncation point to $(-\infty,b_k)$ such that $P_r\{X_k < a_k^*\}$ is negligible. Similarly, if the condition $X_0 = 0$ in (2.3.1) is replaced by the stationary condition $X_0 \sim N(0,(1-\phi^2)^{-1})$, then $\underset{\sim}{\Sigma}$, the variance-covariance matrix of $\underset{\sim}{X}$ becomes

$$\frac{1}{1-\phi^2}\begin{pmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{k-1} \\ \phi & 1 & \phi & \cdots & \phi^{k-2} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \phi^{k-1} & \phi^{k-2} & & \cdots & 1 \end{pmatrix}$$

and

$$\underset{\sim}{\Sigma}^{-1} = \begin{pmatrix} 1 & -\phi & 0 & \cdots & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & \cdots & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & -\phi \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{pmatrix} .$$

Thus, the term $(\sum_{i=1}^{k-1} X_i(\sigma^{ik}+\sigma^{ki})+2X_k\sigma^{kk})$ in (2.3.8) is equal to $-2\phi X_{k-1}+2X_k$ which is the same as in the conditional case. This implies that Theorem 2.3.1 remains valid in the stationary case and since the proofs of both Theorem 2.3.2 and 2.3.4 are independent of the stationarity of $\underset{\sim}{X}$ and the distribution function of X_1, these two theorems still hold even though the condition $X_0 = 0$ in (2.3.1) is removed. Hence, the method proposed in this section can still be used to find the coverage probability for model (2.3.1) when the condition $X_0 = 0$ is replaced by the stationary condition $X_0 \sim N(0,(1-\phi^2)^{-1})$.

CHAPTER III
THE COVERAGE PROBABILITY OF PREDICTION INTERVALS IN AR(1) MODEL

3.1 Computation and Simulation Results

In this section, we first compute the coverage probability of 95% prediction intervals for a first order autoregressive process given $X_0 = 0$. Using the standard prediction method in time series [e.g., Box and Jenkins (1970, Chapter 5)] gives $B_i = (-1.96\sigma_i, 1.96\sigma_i)$ where $\sigma_i^2 = (1+\phi^2+\ldots+\phi^{2(i-1)})$. The values of p_k computed by the algorithm proposed in the last section of the previous chapter from various $\phi = r/s$, n, and k are given in Table II. The values in the parentheses are the same p_k estimated by 10,000 simulations for each case. The simulation was performed on an IBM 4341 computer with its normal random number generator.

From Table II, one can see that the coverage probabilities for various ϕ's do not differ very much for small k, but the gaps are widened when k becomes large. Thus, caution should be employed when these prediction intervals are used for model validation.

To evaluate the error bound, we take the case $\phi = 0.5$ for example. According to Theorems 2.3.1 and 2.3.2, error bound for $k = 10, n = 11$ is

$$|e_k| \leq 55e_0 + [(s/\pi r) + 0.302]\bar{B}_k r^{-11} ,$$

Table II
Coverage Probability of k 95% Confidence Intervals

ϕ	(r,s)	n	k = 3	4	5	10	15	20	25
0.0	-	-	0.8574	0.8145	0.7738	0.5987	0.4633	0.3585	0.2774
0.1	(10,1)	3	0.8572 (0.8680)	0.8141 (0.8193)	0.7731 (0.7811)	0.5973 (0.6056)	0.4615 (0.4695)	0.3566 (0.3642)	0.2755 (0.2840)
0.1667	(6,1)	4	0.8593 (0.8671)	0.8170 (0.8157)	0.7769 (0.7791)	0.6038 (0.6070)	0.4693 (0.4632)	0.3648 (0.3691)	0.2835 (0.2895)
0.20	(5,1)	4	0.8593 (0.8575)	0.8176 (0.8158)	0.7780 (0.7783)	0.6068 (0.6065)	0.4733 (0.4726)	0.3692 (0.3768)	0.2880 (0.2969)
0.25	(4,1)	4	0.8624 (0.8587)	0.8207 (0.8224)	0.7811 (0.7838)	0.6098 (0.6070)	0.4760 (0.4733)	0.3716 (0.3772)	0.2901 (0.2919)
0.30	(10,3)	3	0.8614 (0.8637)	0.8207 (0.8233)	0.7819 (.7874)	0.6134 (0.6087)	0.4813 (0.4908)	0.3776 (0.3840)	0.2962 (0.3013)
0.333	(3,1)	5	0.8639 (0.8742)	0.8240 (0.8260)	0.7856 (0.7885)	0.6187 (0.6199)	0.4873 (0.4900)	0.3838 (0.3866)	0.3023 (0.3125)
0.50	(2,1)	11	0.8693 (0.8722)	0.8323 (0.8356)	0.7972 (0.8001)	0.6424 (0.6517)	0.5176 (0.5241)	0.4170 (0.4204)	0.3360 (0.3461)
0.60	(5,3)	4	0.8736 (0.8801)	0.8399 (0.8394)	0.8075 (0.8097)	0.6648 (0.6647)	0.5267 (0.5385)	0.4509 (0.4558)	0.3714 (0.3689)
0.70	(10,7)	3	0.8784 (0.8835)	0.8478 (0.8470)	0.8193 (0.8239)	0.6926 (0.6866)	0.5861 (0.5881)	0.4960 (0.5035)	0.4197 (0.4287)
0.75	(4,3)	4	0.8816 (0.8811)	0.8528 (0.8523)	0.8260 (0.8277)	0.7097 (0.7125)	0.6107 (0.6067)	0.5254 (0.5343)	0.4520 (0.4527)
0.80	(5,4)	4	0.8832 (0.8833)	0.8566 (0.8586)	0.8323 (0.8325)	0.7277 (0.7261)	0.6383 (0.64134)	0.5606 (0.5675)	0.4923 (0.4925)
0.90	(10,9)	3	0.8878 (0.8905)	0.8651 (0.8605)	0.8454 (0.8487)	0.7667 (0.7683)	0.7024 (0.7056)	0.6453 (0.6487)	0.5933 (0.6008)

where s=1, r=2, $\bar{B}_k = 2x1.96x(1+\phi^2+...+\phi^{18})^{1/2}$, $|e_0| \leq 0.4r^{-11}$, or $|e_k| \leq 0.011$. The simulation error, when evaluated by two standard deviations of the estimate from 10,000 independent Bernoulli trials, is less than 0.01. Thus, it is apparent that the real error e_k propagates much slower than $O(k^2)$ as stated in Theorem 2.3.2. Since most of the coverage probabilities for $k \leq 10$ by our algorithm and the simulation differ much less than 0.01, it seems that the error bounds are overestimated and the actual errors by our algorithm are actually very small. Unfortunately, we were unable to improve our error bounds.

3.2 Applications

Table II can also be used to investigate the change of the coverage probability when the observation gap is changed. It is well known that if X_t is a AR(1) process, the $Y_t = X_{2t}$ is again an AR(1) process with $Y_t - \phi^2 Y_{t-1} = \varepsilon_{2t} + \phi\varepsilon_{2t-1}$. One would expect that for a highly correlated process, $C_y(k) = P_r\{Y_t \varepsilon B^*_{2t}, t=1,2,...,k\}$ can be used as an approximation to $C_x(k) = P_r\{Y_t \varepsilon B_{2t}, t=1,2,...,k\}$, where B^*_{2t} are the 95% prediction intervals for Y_t. This comparison would be useful in quality control when the qualities of items are highly correlated in a product line. Since the coverage probability of prediction intervals at a fixed confidence level is independent of the variance of the white noise process in an AR(1) model, several $C_y(k)$ and $C_x(k)$ can be compared by using Table II. For example, for $\phi = 0.9$, $\phi^2 \doteq 0.8$, $C_y(10) = 0.73$ and $C_x(10) = 0.65$, and for $\phi = 0.7$, $\phi^2 \doteq 0.5$,

$C_y(10) = 0.64$ and $C_x(10) = 0.50$. These gaps seem quite substantial even for $\phi = 0.9$.

The present method can also be used to find the first passage time distribution for model (2.3.1). Let p_k and B_t be defined as before and let

$$T = \min \{ t: X_t \notin B_t \} .$$

Then the distribution of T can be obtained by

$$Pr\{T = k\} = p_{k-1} - p_k .$$

Figure 3.1 shows the values of $Pr\{T = k\}$ for $\phi = 0$, 0.5, 0.75, and 0.9 with $B_t = (-1.96\sigma_i, 1.96\sigma_i)$. It is interesting, but not surprising, to see that the first passage time distribution changes its shape from an independent to a dependent process.

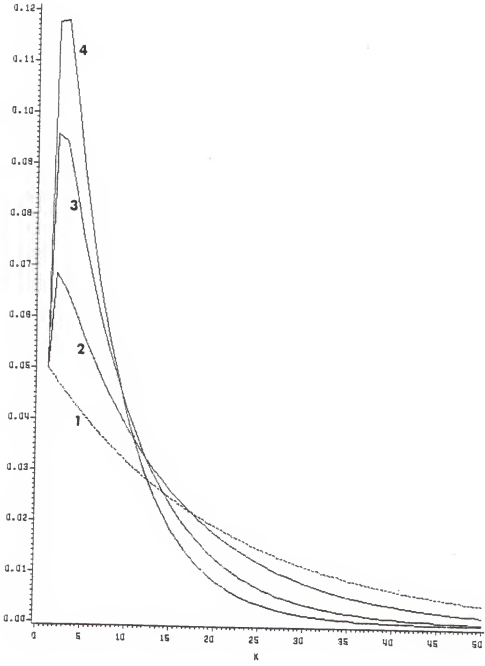


Figure 3.1 First passage time distribution for an AR(1) model. The ϕ values for curves 1, 2, 3, and 4 are respectively 0.0, 0.5, 0.75, and 0.9.

CHAPTER IV
MODEL BUILDING FOR JITTERY COSINE WAVE

4.1 Introduction

Many time series exhibit a variation which is in some way periodic. Some of them have variations at a fixed period due to seasonal effect or due to some other physical cause. For example, the weekly minimum temperature at New Orleans, Louisiana, from 1980 to 1983, whose plot is in Fig. 4.1 presents a variation which is annual in period. Some others exhibit oscillations which do not have a fixed period. The Wolfer's yearly sunspot data from 1749 to 1924 are good examples of this case. The plot in Fig. 4.2 shows that the numbers of years between two local minimum are 11, 9, 9, 14, 12, 13, 10, . . . , and these are quite different from each other.

In analyzing data with variation at a fixed period, taking difference between data to remove the periodicity is one of the most common methods (e.g., Box and Jenkins, 1976, Chapter 9). Decomposing the data into a sum of deterministic periodical functions and a random component is another one (e.g., Anderson, 1971, Chapter 4). Since the first method implicitly requires a fixed period and the second one also explicitly shows that the fixed period is necessary, both of them are inadequate in analyzing cyclical data without a fixed period. Some authors (e.g., Moran, 1954; Box and Jenkins, 1976; Woodward and Gray, 1978) have used the ARMA model to fit the sunspot data. This is

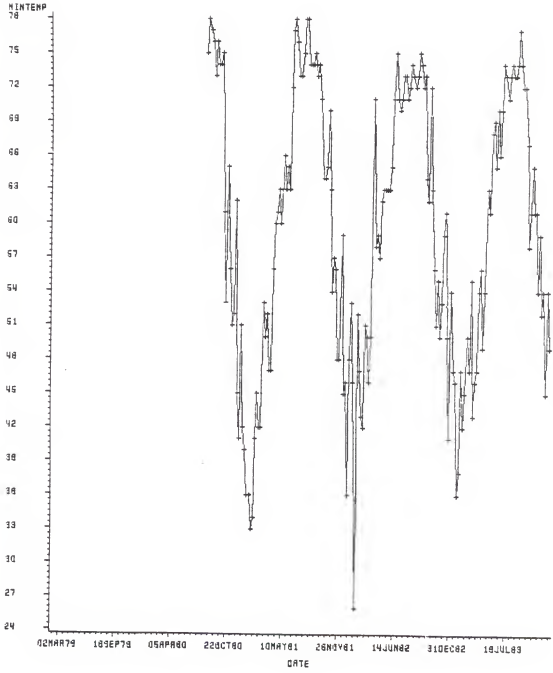


Figure 4.1 Weekly minimum temperature at New Orleans, Louisana.

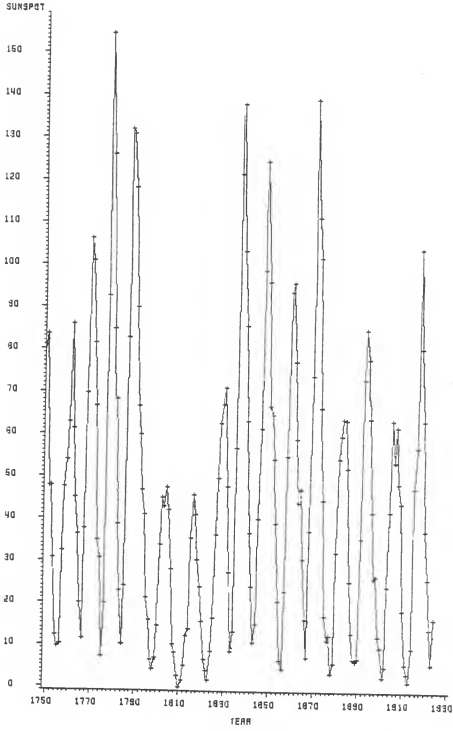


Figure 4.2 Wolfer's yearly sunspot data.

obviously inadequate because they did not incorporate the existence of periodicity in their model.

In this chapter, we start with a simple stationary time series model which is basically a cosine function with jitteriness in frequency ω from time to time. Formal definition of this model follows.

4.2 Description of the Model and Its Stationarity

Let $X_t = a_0 \cos(\omega_0 t + \theta_t) \quad t = 0,1,2, \ldots$ (4.2.1)

where

$a_0 > 0$ and $\omega_0 \in (0,2\pi)$ are real constants,

$\theta_t = \sum_{i=0}^{t} \varepsilon_i$ with

$\varepsilon_0 \sim U(0,2\pi)$ and

$\{\varepsilon_i, i \geq 1\}$ are i.i.d. normal random variables with zero mean and variance σ^2, and ε_0 is independent of $\{\varepsilon_i, i > 1\}$.

Since $\cos(\omega_0 t+\theta_t) = \cos((\omega_0 t+\theta_t) \bmod (2\pi))$, without loss of generality, we can study the distribution of $(\omega_0 t+\theta_t) \bmod (2\pi)$.

To do so, we first give the following lemma.

Lemma 4.2.1 If $Z \sim U(0,2\pi)$, then for any random variables X and integer i, $(iZ+X) \bmod (2\pi) \sim U(0,2\pi)$.

Proof: It is trivial that $(iZ)\mathrm{mod}(2\pi) \sim U(0,2\pi)$ and $(iZ+x)\mathrm{mod}(2\pi) \sim U(0,2\pi)$, where x is a constant.

Let $Y = (iZ+X)\mathrm{mod}(2\pi)$, then

$$\Pr(Y<a) = \Pr((iZ+X)\mathrm{mod}(2\pi)<a)$$

$$= \int_X \Pr((iZ+x)\mathrm{mod}(2\pi)<a)\,dF(x)$$

$$= \Pr((Z)\mathrm{mod}(2\pi)<a) \int_X dF(x)$$

$$= \Pr((Z)\mathrm{mod}(2\pi)<a).$$

From Lemma 4.2.1, we have

$$(\omega_0 t+\theta_t)\mathrm{mod}(2\pi) \sim U(0,2\pi) \qquad t=1,2,\ldots \ .$$

Moreover,

$$(\omega_0 i+\theta_j)\mathrm{mod}(2\pi) \sim U(0,2\pi) \qquad i,j=1,2,\ldots \ . \tag{4.2.2}$$

With this result letting $\Delta_{ij} = \theta_j - \theta_i = \sum_{K=i+1}^{j} \varepsilon_K$, we can derive some basic properties of $\{X_t\}$ defined in (4.2.1) as follows. Throughout the rest of this chapter, let Z denote a random variable uniformly distributed on $(0,2\pi)$.

Stationarity

Theorem 4.2.2 The process $\{X_t\}$ defined in (4.2.1) is a stationary time series.

Proof: It suffices to show that

$$\Pr(X_i \le a_1,\ X_{i+n_1} \le a_2, \ldots, X_{i+n_m} \le a_{m+1})$$

$$= \Pr(X_{i+t} \le a_1,\ X_{i+n_1+t} \le a_2, \ldots,\ X_{i+n_m+t} < a_{m+1})$$

where i, $n_1 < n_2 < \ldots < n_m$ and m are positive integers and $a_1, a_2, \ldots, a_{m+1}$ are real constants.

Since

$$\Pr(X_i \le a_1,\ X_{i+n_1} \le a_2, \ldots,\ X_{i+n_m} \le a_{m+1})$$

$$= \Pr\{a_0\cos(i\omega_0+\theta_i) \le a_1, a_0\cos((i+n_1)\omega_0+\theta_i+\Delta_{i,i+n_1}) \le a_2,$$

$$\ldots,\ a_0\cos((i+n_m)\omega_0+\theta_i+\Delta_{i,i+n_m}) \le a_{m+1}\}$$

$$= \Pr\{a_0\cos((i\omega_0+\theta_i)\bmod(2\pi)) \le a_1,$$

$$a_0\cos((i\omega_0+\theta_i)\bmod(2\pi)+n_1\omega_0+\Delta_{i,i+n_1}) \le a_2,$$

$$\ldots,\ a_0\cos((i\omega_0+\theta_i)\bmod(2\pi)+n_m\omega_0+\Delta_{i,i+n_m}) \le a_{m+1}\}$$

$$= \Pr\{a_0\cos(Z) \le a_1,\ a_0\cos(Z+n_1\omega_0+\Delta_{i,i+n_1}) \le a_2, \ldots$$

$$a_0\cos(Z+n_m\omega_0+\Delta_{i,i+n_m}) \le a_{m+1}\}\ .$$

Similarly,

$$\Pr\{X_{i+t} \le a_1,\ X_{i+t+n_1} \le a_2, \ldots,\ X_{i+t+n_m} \le a_{m+1}\}$$

$$= \Pr\{a_0\cos(Z) \le a_1,\ a_0\cos(Z+n_1\omega_0+\Delta_{i+t,i+t+n_1}) \le a_2,$$

$$\ldots,\ a_0\cos(Z+n_m\omega_0+\Delta_{i+t,i+t+n_m}) \le a_{m+1}\}$$

$$= \Pr\{a_0\cos(Z) \le a_1,\ a_0\cos(Z+n_1\omega_0+\Delta_{i,i+n_1}) \le a_2,$$

$$\ldots,\ a_0\cos(Z+n_m\omega_0+\Delta_{i,i+n_m}) \le a_{m+1}\} .$$

Hence, the process $\{X_t\}$ defined in (4.2.1) is a stationary process.

Moments and Autocorrelation Function

The mean is

$$EX_t = Ea_0\cos(\omega_0 t+\theta_t) = Ea_0\cos(Z) = 0 \qquad t=1,2,\ldots .$$

The autocovariance function is

$$r(K) = E\{X_tX_{t+K}\} - (EX_t)^2$$

$$= E\{a_0\cos(\omega_0 t+\theta_t)\ a_0\cos(\omega_0(t+K)+\theta_{t+K}\} - 0$$

$$= a_0^2E\{\cos(Z)+\cos(K\omega_0+\Delta_{t,t+K})\}/2$$

$$= (a_0^2\cos(\omega_0K)e^{-K\sigma^2/2})/2 \qquad K=0,1,2,\ldots .$$

Then,

$$\mathrm{var}(X_t) = r(0) = a_0^2/2 .$$

So, the autocorrelation function is

$$\rho(K) = (EX_tX_{t+K})/\text{var}(X_t) = \cos(\omega_0 K)e^{-K\sigma^2/2} .$$

The expectation of the third order cross product is

$$\begin{aligned} EX_tX_{t+\ell}X_{t+m} &= a_0^3 E\{\cos(\omega_0 t+\theta_t)\cos(\omega_0(t+\ell)+\theta_{t+\ell}) \cdot \\ &\qquad \cos(\omega_0(t+m)+\theta_{t+m})\} \\ &= a_0^3 E\{\cos(3t+\ell+m)\omega_0+\theta_t+\theta_{t+\ell}+\theta_{t+m}) \\ &\qquad +\cos((t+\ell-m)\omega_0+\theta_t+\theta_{t+\ell}-\theta_{t+m}) \\ &\qquad +\cos((t+\ell+m)\omega_0+\theta_{t+\ell}-\theta_t+\theta_{t+m}) \\ &\qquad +\cos((t-\ell+m)\omega_0+\theta_{t+m}-\theta_{t+\ell}-\theta_t)\}/4 \\ &= 0 . \end{aligned}$$

The Spectrum Function

By definition,

$$\begin{aligned} f(\lambda) &= 2(1+2\sum_{k=1}^{\infty}\rho(k)\cos\lambda k) \qquad 0 \le \lambda \le \pi \\ &= 2+4\sum_{k=1}^{\infty}\cos(\omega_0 k)\cos(\lambda k)e^{-k\sigma^2/2} \\ &= 2+2\sum_{k=1}^{\infty}e^{-k\sigma^2/2}[\cos((\omega_0+\lambda)k)+\cos((\omega_0-\lambda)k)], \end{aligned}$$

and since

$$\sum_{k=1}^{n-1} \rho^k \cos kx = \frac{1-\rho\cos x-\rho^n\cos nx+\rho^{n+1}\cos(n-1)x}{1-2\rho\cos x+\rho^2} \to \frac{1-\rho\cos x}{1-2\rho\cos x+\rho^2} \quad \text{as } n\to\infty$$

we have

$$f(\lambda) = 2+2[(1-e^{-\sigma^2/2}\cos(\omega_0+\lambda))/(1-2e^{-\sigma^2/2}\cos(\omega_0+\lambda)+e^{-\sigma^2}) +$$

$$(1-e^{-\sigma^2/2}\cos(\omega_0-\lambda))/(1-2e^{-\sigma^2/2}\cos(\omega_0-\lambda)+e^{-\sigma^2})] . \qquad (4.2.3)$$

It is well known (see e.g., Chatfield, 1980, Chapter 4) that the spectrum function of a deterministic sinusoidal perturbation which has a definition similar to (4.2.1) except that

$$\theta_i = \varepsilon_0 \quad \text{for any } i \geq 1 ,$$

has its maximum value at $\lambda = \omega_0$. But this is not true in our model because

$$\left.\frac{\partial f(\lambda)}{\partial \lambda}\right|_{\lambda=\omega_0} = e^{-\sigma^2/2}\sin 2\omega_0(e^{-\sigma^2}-1) > 0 \quad \text{for } \omega_0 \neq 0 .$$

To find the λ value which maximizes $f(\lambda)$ is quite difficult. Instead of doing so, we give two plots, Fig. 4.3 and Fig. 4.4., to illustrate some behaviors of $f(\lambda)$.

Figure 4.3 is a plot of $f(\lambda)$ when $\sigma^2 = 1.0$ and $\omega_0 = 0.4$, from which one can see that $f(\lambda)$ does not have the maximum value at $\lambda = \omega_0$. Figure 4.4 is a plot of $\omega^*(\sigma^2,\omega_0)$ when $\omega_0 = 0.4$, where $\omega^*(\sigma^2,\omega_0)$ is the λ value which maximizes $f(\lambda)$. From this plot, one

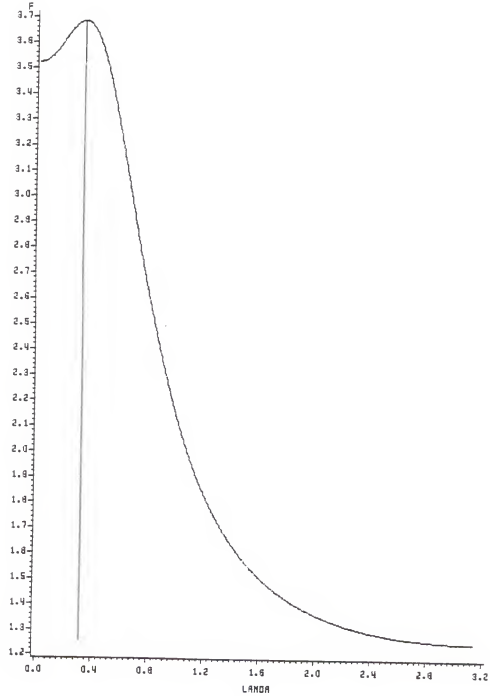


Figure 4.3 Spectrum density of model (4.2.1) when $\sigma^2 = 1.0$ and $\omega_0 = 0.4$.

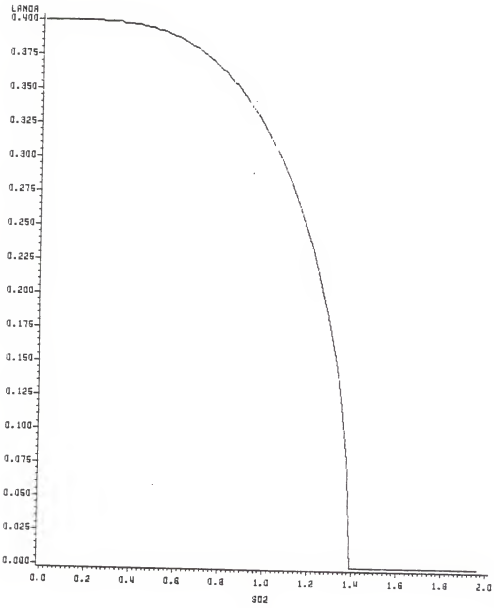


Figure 4.4 Plot of $\omega^*(\sigma^2, 0.4)$ vs σ^2

can see how $\omega^*(\sigma^2,0.4)$ departs from ω_0 when σ^2 gets larger. Since $f(\lambda)$ does not have its maximum value at $\lambda = \omega_0$, the classical spectrum analysis cannot be used here.

4.3 Minimum Mean Square Error Predictor

By a well-known result, the mean square error of the ℓ-step predictor of $X_{t+\ell}$ at time t, given the knowledge of the model and all the X's up to t, is minimized when it is equal to the conditional expected value of $X_{t+\ell}$. Thus, the ℓ-step minimum mean square error predictor at time t is

$$\hat{X}_t(\ell) = E\{X_{t+\ell}|a_0,\omega_0,\sigma^2,X_t,X_{t-1},\ldots\}$$

$$= E\{a_0\cos((\omega_0(t+\ell)+\theta_t+\Delta_{t,t+\ell})|\omega_0,a_0,\theta_t,\sigma^2)\}$$

$$= a_0 e^{-\ell\sigma^2/2}\cos(\omega_0(t+\ell)+\theta_t) \ .$$

Let $e_t(\ell)$ denotes the ℓ-step prediction error at time t, then

$$e_t(\ell) = X_{t+\ell} - a_0 e^{-\ell\sigma^2/2}\cos(\omega_0(t+\ell)+\theta_t) \ ,$$

and the variance of $e_t(\ell)$ denoted as $V(e_t(\ell))$ is

$$V(e_t(\ell)) = E\{X_{t+\ell} - a_0 e^{-\ell\sigma^2/2}\cos(\omega_0(t+\ell)+\theta_t)\}^2$$

$$= E\{(X_{t+\ell})^2\} - a_0^2 e^{-\ell\sigma^2}\cos^2(\omega_0(t+\ell)+\theta_t)$$

$$= (a_0^2/2)(\cos(2\omega_0(t+\ell)+2\theta_t)e^{-2\ell\sigma^2}+1) - a_0^2e^{-\ell\sigma^2}\cos^2(\omega_0(t+\ell)+\theta_t)$$

$$= a_0^2\cos^2(2(t+\ell)\omega_0+2\theta_t)e^{-2\ell\sigma^2} - a_0^2e^{-\ell\sigma^2}/2 + a_0^2/2 -$$

$$a_0^2e^{-\ell\sigma^2}\cos(2(t+\ell)\omega_0+2\theta_t$$

$$= a_0^2(1-e^{-\ell\sigma^2})(e^{-\ell\sigma^2}(1/2 - \cos^2((t+\ell)\omega_0+\theta_t))+1/2)$$

$$= a_0^2(1-e^{-\ell\sigma^2})(-e^{-\ell\sigma^2}\cos(2(t+\ell)\omega_0+2\theta_t)+1)/2 \ . \qquad (4.3.1)$$

In the ARMA(p,q) process $V(e_t(\ell))$ goes to the variance of X when ℓ goes to infinity. This is also true in the process defined by (4.2.1). But as can be seen in (4.3.1), $V(e_t(\ell))$ has maximum value while $\cos((t+\ell)\omega_0+\theta_t) = 0$ and minimum value while $\cos((t+\ell)\omega_0+\theta_t) = \pm 1$. So $V(e_t(\ell))$ depends on t while it is independent of t in ARMA(p,q) process. This is due to the fact that the change in cosine wave is small when its angle is close to 0 or π and is large when its angle is close to $\pi/2$.

Another difference about $V(e_t(\ell))$ between these two processes is that in the ARMA(p,q) process $V(e_t(\ell))$ increases when ℓ increases, but this is not necessarily true in the process defined in (4.2.1). In order to illustrate how $V(e_t(\ell))$ behaves when ℓ increases while t, ω_0 and σ^2 are fixed, let us assume that $\pi = \omega_0 k$ where k is some positive integer, and let $A_{t,\ell}$ denote $\cos(2(t+\ell)\omega_0+2\theta_t)$. Then

$\forall\ \ell' = \ell + mk \qquad m = 1,2,\ldots,$

$V(e_t(\ell')) - V(e_t(\ell))$

$= a_0^2\{e^{-\ell\sigma^2} - e^{-(\ell+mK)\sigma^2} - A_{t,\ell}e^{-\ell\sigma^2}(e^{-mK\sigma^2} + e^{-(\ell+mK)\sigma^2} - 1 + e^{-\ell\sigma^2})\}/2$

$= a_0^2\{e^{-\ell\sigma^2}(1-e^{-mK\sigma^2})(1-A_{t,\ell}(e^{-\ell\sigma^2}(1+e^{-mK\sigma^2}) - 1))\}/2\ ,$

which is greater than zero, because

$-1 \le e^{-\ell\sigma^2}(1+e^{-m\sigma^2}) - 1 \le 1$ and

$-1 \le A_{t,\ell} \le 1\ .$

That is, $V(e_t(\ell))$ is monotone increasing from period to period.

In the case of $\ell' = \ell + j \qquad j=1,2,\ldots,k-1$:

(i) If $A_{t,\ell'} \le A_{t,\ell}$ then

$$(1-e^{-\ell'\sigma^2})(1-e^{-\ell'\sigma^2}A_{t,\ell'}) > (1-e^{-\ell'\sigma^2})(1-e^{-\ell'\sigma^2}A_{t,\ell})\ .$$

Hence,

$$V(e_t(\ell')) \ge V(e_t(\ell))\ . \qquad (4.3.2)$$

(ii) If $A_{t,\ell'} > A_{t,\ell}$ then (4.3.2) might not hold; for example, when $A_{t,\ell'} = 1$ and $A_{t,\ell'} = -1$ then $V(e_t(\ell')) - V(e_t(\ell))$

$= (1-e^{-(\ell+K)\sigma^2})(1-e^{-(\ell+K)\sigma^2}) - (1-e^{-\ell\sigma^2})(1+e^{-\ell\sigma^2})$

$$= e^{-\ell\sigma^2}(e^{-(2K+\ell)\sigma^2} - 2e^{-K\sigma^2} + e^{-\ell\sigma^2})$$

$$< e^{-\ell\sigma^2}(2e^{-\ell\sigma^2} - 2e^{-K\sigma^2})$$

which is less than zero if $k < \ell$. Therefore, function $V(e_t(\ell))$ is no longer monotone increasing in ℓ within period.

Some plots of $V(e_t(\ell))$ shown in Fig. 4.5 give a clearer picture of what $V(e_t(\ell))$ would look like.

In Fig. 4.5,

line 4 is the plot of $V(e_t(\ell))+3$ computed when

$$a_0^2 = 2,\ \sigma^2 = 0.1,\ \omega_0 = \pi/20 \text{ and } 2\theta_t + 2t\omega_0 = 0.2,$$

line 3 is the plot of $V(e_t(\ell))+2$ computed when

$$a_0^2 = 2,\ \sigma^2 = 0.1,\ \omega_0 = \pi/10 \text{ and } 2\theta_t + 2t\omega_0 = 0.2,$$

line 2 is the plot of $V(e_t(\ell))+1$ computed when

$$a_0^2 = 2,\ \sigma^2 = 0.1,\ \omega_0 = \pi/20 \text{ and } 2\theta_t + 2t\omega_0 = 0$$

and

line 1 is the plot of $V(e_t(\ell))$ computed when

$$a_0^2 = 2,\ \sigma^2 = 0.1,\ \omega_0 = \pi/10 \text{ and } 2\theta_t + 2t\omega_0 = 0\ .$$

4.4 A Special Case

While it is true that a given model has a unique autocovariance structure, the converse is not true. In this section, we give a

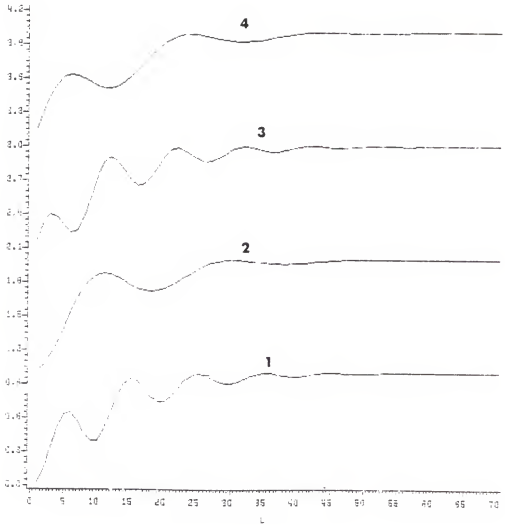


Figure 4.5 Plot of $V(e_t(\ell))$ for model (4.2.1).

special case of model (4.2.1) which has the same autocovariance as that of an AR(1) model.

Let $\{Y_t\}$ be a process defined by (4.2.1) with $\omega_0 = 0$ and $\{X_t\}$ be an AR(1) process defined as follows:

$$X_t = \phi X_{t-1} + E_t \quad , \text{ where}$$

$$|\phi| < 1 \qquad \text{and}$$

$\{E_t\}$ are i.i.d. normal random variables with zero mean and variance σ_E^2.

If $\phi = e^{-\sigma^2/2}$ and $\sigma_E^2 = a_0^2(1-e^{-\sigma^2})/2$, then $\{Y_t\}$ and $\{X_t\}$ have the same autocovariance structure; that is,

$$\rho_X(k) = \phi^k = (e^{-\sigma^2/2})^k = \rho_Y(k) \qquad \text{and}$$

$$\text{var}(X_t) = \frac{\sigma_E^2}{(1-\phi^2)} = \frac{a_0^2}{2} = \text{var}(Y_t) \ .$$

The minimum mean square error ℓ-step predictor of $X_{t+\ell}$ is

$$E(X_{t+\ell} | \phi, X_t, X_{t-1}, \ldots) = \phi^\ell X_t = \hat{X}_t(\ell)$$

which has the same form as the minimum mean square error ℓ-step predictor of $Y_{t+\ell}$ because

$$\hat{Y}_t(\ell) = a_0 e^{-\ell\sigma^2/2}\cos(\theta_t) = e^{-\ell\sigma^2/2} Y_t = \phi^\ell Y_t \ .$$

The variances of the ℓ-step prediction error for these models are

$$\mathrm{var}(X_{t+\ell}-\hat{X}_t(\ell)) = \sum_{i=0}^{\ell-1} \phi^{2i} \sigma_\varepsilon^2 = a_0^2(1-e^{-\ell\sigma^2})/2$$

and

$$\mathrm{var}(Y_{t+\ell}-\hat{Y}_t(\ell)) = a_0^2(1-e^{-\ell\sigma^2})(1-\cos(2\theta_t)e^{-\ell\sigma^2})/2 \quad .$$

In addition,

$$\mathrm{var}(Y_{t+\ell}-\hat{Y}_t(\ell)) > \mathrm{var}(X_{t+\ell}-\hat{X}_t(\ell)) \text{ when}$$

$$1-\cos(2\theta_t)e^{-\ell\sigma^2} > 1$$

$$\Leftrightarrow \cos(2\theta_t) > 0$$

$$\Leftrightarrow 2\cos^2(\theta_t) - 1 > 0$$

$$\Leftrightarrow Y_t > a_0/\sqrt{2} \text{ or } Y_t < -a_0/\sqrt{2}$$

and

$$\Pr(|Y_t|>a_0/\sqrt{2}) = \Pr(|\cos(\omega_0 t+\theta_t)|>1/\sqrt{2}) = 1/2 \quad .$$

Hence, by the criterion of the variance of the ℓ-step prediction error, we cannot say exactly which model is better.

4.5 Consistent Estimators

It would be desirable to find consistent estimators for the parameters of model (4.2.1) by utilizing all available data. But we

are unable to do so. However, we find a consistent estimator for the parameter a_0 by using all observations and a consistent estimator for the parameter ω_0 by using a proportion of the observations.

Lemma 4.5.1 Let $\alpha_n = \inf_{1\le t\le n} ((\omega_0 t+\theta_t) \bmod (2\pi))$, then

$$\alpha_n \to 0 \quad \text{a.s.} \quad \text{as } n\to\infty .$$

Proof: Since $\{(\omega_0 t+\theta_t) \bmod (2\pi),\ t\ge 0\}$ is a random walk on a circle and ε_0 is the starting point of the process, without loss of generality, we assume that $\varepsilon_0 = 0$ and write

$$S_n = (\sum_{i=1}^{n} \varepsilon_i') \bmod (2\pi) \qquad \text{where } \varepsilon_i' = \varepsilon_i + \omega_0 ,$$

then

$$\Pr\{|S_n| < \delta\} > \int_0^{\delta} e^{-\frac{(X-\omega_0 n)^2}{2n\sigma^2}} / ((2\pi)^{1/2}\sigma)\, dx > C_t/n^{1/2}$$

where $C_t > 0\ \forall\ t$.

Thus, we have

$$\sum_{n=1}^{\infty} \Pr\{|S_n| < \delta\} = \infty .$$

Then following the proof given by Chow and Teicher (1978, example 4.2.1), we can prove that

$$P(|S_n| < \delta \text{ i.o.}(n)) = 1 \qquad \forall\ \delta > 0 .$$

This implies that

$$\liminf_{n\to\infty} S_n = 0 \qquad \text{a.s.} \ .$$

Thus $\alpha_n \to 0$ a.s. as $n\to\infty$.

Lemma 4.5.2 Let $\hat{a}_n = \max(|X_1|, |X_2|, \ldots, |X_n|)$, then

$$\hat{a}_n \to a_0 \quad \text{a.s.} \qquad \text{as } n \to \infty \ .$$

Proof: If $B_1 = \{\omega: \lim_{n\to\infty} \hat{a}_n = a_0\}$, $B_2 = \{\omega: \lim_{n\to\infty} \alpha_n = 0\}$,

then $B_2 \subseteq B_1$. So, by lemma 4.5.1,

$$\Pr(B_1) \geq \Pr(B_2) = 1 \ .$$

Hence, by lemma 4.5.2, $\hat{a}_n$ is a consistent estimator for parameter a_0.

Before giving a consistent estimator for ω_0, we state a theorem which was given by C.F. Wu in 1981.

Theorem 4.5.3

Let $\hat{\theta}_n$ be an estimator of θ_0 which is based on the minimization or maximization of a function $S_n(\theta)$. Then for any $\delta > 0$, if

$$\liminf_{n\to\infty} \inf_{|\theta-\theta_0|>\delta} (S_n(\theta)-S_n(\theta_0)) > 0 \quad \text{a.s. (in prob)}$$

then $\hat{\theta}_n \to \theta_0$ a.s. (in prob) as $n\to\infty$.

Since $\{\epsilon_t \ t\geq 1\}$ appears in model (4.2.1) only through the cosine

function which is a periodical function of period 2π, without loss of generality, we condense all the probability of ε_t in the real line to $(-\pi,\pi)$; that is, the density function $f_t(x)$ of the condensed variable ε_t^* is

$$f_t(x) = \sum_{K=-\infty}^{\infty} g_t(x+2\pi K) \qquad -\pi < x \leq \pi\ , \quad t=1,2,\ldots,$$

where $g_t(x)$ is the density function of ε_t.

Then the process $\{X_t\}$ defined in (4.2.1) can be redefined by substituting $\{\varepsilon_t,\ t>1\}$ by $\{\varepsilon_t^*,\ t>1\}$. Since

$$(\theta_t+\omega_0 t)\bmod(2\pi) = \cos^{-1}\frac{X_t}{a_0} \text{ or } 2\pi-\cos^{-1}\frac{X_t}{a_0}\ , \text{ and}$$

$$(\theta_{t-1}+\omega_0(t-1))\bmod(2\pi) = \cos^{-1}\frac{X_{t-1}}{a_0} \text{ or } 2\pi-\cos^{-1}\frac{X_{t-1}}{a_0}\ ,$$

$\varepsilon_t^*+\omega_0$ may be equal to one of the following values:

$$\cos^{-1}(X_t/a_0) - \cos^{-1}(X_{t-1}/a_0) \quad \text{or}$$

$$2\pi - \cos^{-1}(X_t/a_0) - \cos^{-1}(X_{t-1}/a_0) \quad \text{or}$$

$$2\pi - \cos^{-1}(X_t/a_0) - (2\pi - \cos^{-1}(X_{t-1}/a_0) \quad \text{or}$$

$$\cos^{-1}(X_t/a_0) - (2\pi - \cos^{-1}(X_{t-1}/a_0))\ .$$

Thus, we can write

$$\varepsilon_t^*+\omega_0 = I_{1t}(\cos^{-1}X_t/a_0-\cos^{-1}X_{t-1}/a_0)+I_{2t}(2\pi-\cos^{-1}X_t/a_0-\cos^{-1}X_{t-1}/a_0)$$

$$+I_{3t}(\cos^{-1}X_{t-1}/a_0-\cos^{-1}X_t/a_0)+I_{4t}(\cos^{-1}X_t/a_0+\cos^{-1}X_{t-1}/a_0-2\pi) \qquad (4.5.1)$$

where

$\{I_{it},\ i=1,2,3,4\}$ are zero-one random variables whose values depend on $\omega_0 t+\theta_t$ and $\omega_0(t-1)+\theta_{t-1}$, and only one of I_{it}'s is equal to 1 at each time t.

To find a consistent estimator of ω_0 through (4.5.1) without knowing the value of I_{it}'s is very difficult. Instead of doing this, we would like to find a consistent estimator for ω_0 conditional on the knowledge of $\{I_{it},\ t\geq 1\}$.

But it is impossible to identify I_{it}'s for all t, especially when either X_t or X_{t-1} is near by a_0 or $-a_0$. Thus, we give an estimator of ω_0 by using only those X_t's of which I_{it} can be identified. And it is possible to identify I_{it} for some t when both σ^2 and ω_0 are relatively smaller than π; for example, if $\sigma^2 = \omega_0 = 0.1\pi$ then

$$\Pr(|\varepsilon_t^*+\omega_0|>0.5\pi) \sim \Pr(|\varepsilon_t+\omega_0|>0.5\pi) = 1-\Phi(4)+\Phi(-6) = 0.00003 .$$

Thus

$$\Pr(I_{2t}=1 \text{ or } I_{4t}=1 \mid |\cos^{-1}X_t/a_0-\cos^{-1}X_{t-1}/a_0|<0.5\pi) = 0.00003$$

and either $I_{1t} = 1$ or $I_{3t} = 1$ can be determined by the direction of the movement of data around t.

Let $T = \{t: 1 \leq t \leq n \text{ where } I_{it} \text{ are known}\}$, and let n_T be the number of points in set T. And for fixed n, n_T decreases when σ^2 or ω_0 increase. Now an estimator of ω_0 based only on those X_i's where $i \in T$ is given as follows:

$$\hat{\omega}_{n_T}(a_0) = \sum_{i \in T} (\epsilon_i^* + \omega_0)/n_T$$

$$= \sum_{i \in T} f_i(X_i, X_{i-1}, a_0)/n_T, \text{ where}$$

$f_i(X_i, X_{i-1}, a_0) = \cos^{-1}X_i/a_0 - \cos^{-1}X_{i-1}/a_0$ or $2\pi - \cos^{-1}X_i/a_0 - \cos^{-1}X_{i-1}/a_0$

or

$\cos^{-1}X_{i-1}/a_0 - \cos^{-1}X_i/a_0$ or $\cos^{-1}X_i/a_0 + \cos^{-1}X_{i-1}/a_0 - 2\pi$.

Theorem 4.5.4

$\hat{\omega}_{n_T}(a_0)$ is a consistent estimator of ω_0.

Proof: It can be seen that $\hat{\omega}_{n_T}(a_0)$ is an estimator which minimizes

$$S_n(\omega) = \sum_{i \in T} (\epsilon_i^*)^2 = \sum_{i \in T} (f_i(X_i, X_{i-1}, a_0) - \omega)^2 .$$

And

$$S_n(\omega) - S_n(\omega_0) = -\sum_{i \in T} \epsilon_i^{*2} + \sum_{i \in T} (\epsilon_i^{*2} + \omega_0 - \omega)$$

$$= \sum_{i \in T} (\omega_0 - \omega)^2 - 2(\omega_0 - \omega) \sum_{i \in T} \epsilon_i^{*2}$$

and from the definitions of ε_i and ε_i^*, we have

$$E\ \varepsilon_i^* = 0,\ \mathrm{var}(\varepsilon_i^*) < \sigma^2 < \infty .$$

Thus, we have $\dfrac{S_n(\omega)-S_n(\omega_0)}{n_T} = (\omega_0-\omega)^2 - \dfrac{2(\omega_0-\omega)\ \sum_{i\varepsilon T} \varepsilon_i^*}{n_T} > 0$ a.s. for any $\omega \neq \omega_0$. Then, by Theorem 4.5.3, we can conclude that $\hat{\omega}_{n_T}(a_0)$ is a consistent estimator of ω_0.

We now prove that $\hat{\omega}_{n_T}(\hat{a}_n)$ is a consistent estimator of ω_0 when all X_i, $i\varepsilon T$, are bounded away from a_0. The Taylor expansion of $\hat{w}_{n_T}(\hat{a}_n)$ at a_0 is

$$\hat{\omega}_{n_T}(\hat{a}_n) = \hat{\omega}_{n_T}(a_0)+(a_0-\hat{a}_n)\ \frac{\partial}{\partial a}\ \hat{\omega}_{n_T}(a)\Big|_{a=a^*}\ ,\ \text{where } a^*\varepsilon[\hat{a}_n, a_0]$$

$$\frac{\partial}{\partial a}\ \hat{\omega}_{n_T}(a) = \frac{\partial}{\partial a}\ (1/n_T)\ \sum_{i\varepsilon T} f_i(X_i, X_{i-1}, a) .$$

Since $\dfrac{\partial}{\partial a} \cos^{-1} \dfrac{X_t}{a} = -X_t/(a\sqrt{a^2-X_t^2})$, it can be seen that $\dfrac{\partial}{\partial a}\ f_t(X_t, X_{t-1}, a)$ has the form

$$\beta_{1t}X_t/(a\sqrt{a^2-X_t^2})+\beta_{2t}X_{t-1}/(a\sqrt{a^2-X_{t-1}^2})$$

where

$$\beta_{1t} = +1 \text{ or } -1 \text{ and } \beta_{2t} = +1 \text{ or } -1 \text{ depending on } t,$$

and by lemma 4.5.2, $\hat{a}_n \to a_0$ a.s. ,

then we have $a^* \to a_0$ a.s. , because $a^* \in [\hat{a}_0, a_0]$.

These results, together with the assumption that X_t's are bounded away from a_0, imply

$$\frac{1}{a^*\sqrt{a^{*2}-X_t^2}} < \infty \quad \text{and} \quad \frac{1}{a^*\sqrt{a^{*2}-X_{t-1}^2}} < \infty .$$

Thus

$$\frac{\partial}{\partial a} \hat{\omega}_{n_T}(a) \Big|_{a=a^*} \text{ is less than infinity.}$$

Hence

$$(a_0 - \hat{a}_n) \frac{\partial}{\partial a} \hat{\omega}_n(a) \Big|_{a=a^*} \to 0 \quad \text{a.s.}$$

And by theorem 4.5.4, $\hat{\omega}_{n_T}(a_0) \to \omega_0$ a.s. ; therefore, we can conclude that $\hat{\omega}_{n_T}(\hat{a}_n) \to \omega_0$ a.s. That is, $\hat{\omega}_{n_T}(\hat{a}_n)$ is a consistent estimator of ω_0.

CHAPTER V
FUTURE RESEARCH: GENERAL MODEL AND SUNSPOT DATA

5.1 General Model

The model proposed in section 4.2 includes only one frequency and is a rather simple model. To explain some phenomena in the real world, we may need a more complicated model that contains more than one frequency component. A model more general than (4.2.1) can be defined as follows:

$$X_t = \alpha_0 + \sum_{j=1}^{m} a_j \cos(j(\omega_0 t + \theta_t)) \qquad (5.1.1)$$

where

α_0 and a_j's > 0 are real constants, and ω_0, θ_t are the same as those defined in (4.2.1).

Then process $\{X_t\}$ defined in (5.1.1) is again a stationary process. The proof is omitted since it is identical to the proof of the stationarity of model (4.2.1). Similar properties are derived as follows:

The means of X_t is

$$EX_t = \alpha_0 .$$

The autocovariances are

$$r(k) = EX_tX_{t+k} - (EX_t)^2$$

$$= \sum_{j=1}^{m} (a_j^2/2)\cos(jk\omega_0)e^{-j^2k\sigma^2/2} \quad ,$$

$$\mathrm{var}(X_t) = r(0) = \sum_{j=1}^{m} a_j^2/2 \; .$$

The autocorrelation function is

$$\rho(k) = \sum_{j=1}^{m} a_j^2\cos(jk\omega_0)e^{-j^2k\sigma^2/2} \Big/ \sum_{j=1}^{m} a_j^2 \quad .$$

The spectrum function is

$$f(\lambda) = 2+2\sum_{j=1}^{m} f_j(\lambda) \Big/ \sum_{j=1}^{m} a_j^2 \qquad 0 \le \lambda \le \pi \; ,$$

where

$$f_j(\lambda) = a_j^2\{(1-e^{-j^2\sigma^2/2}\cos(j\omega_0+\lambda))/(1-2e^{-j^2\sigma^2/2}\cos(j\omega_0+\lambda)+e^{-j^2\sigma^2})$$

$$+(1-e^{-j^2\sigma^2/2}\cos(j\omega_0-\lambda))/(1-2e^{-j^2\sigma^2/2}\cos(j\omega_0-\lambda)+e^{-j^2\sigma^2})\} \; .$$

But the method of estimating ω_0 and a_j's in section 4.5 cannot be easily extended to model (5.1.1), because it includes more than one a_j and ω_0 cannot be extracted easily by inverting the function $\sum_{j=1}^{m} a_j\cos(j(\omega_0 t+\theta_t))$. Parameter estimation for model (5.1.1) has not been done, but it warrants further research.

5.2 Modeling Sunspot Data

Wolfer's sunspot data have attracted many authors' interest. Woodward and Gray (1978) summarize models proposed by other authors and gave three new models; they are

$$(1-1.64B+B^2)(1-.29B^4-.21B^5-.19B^6)X_t = (1-.34B)a_t, \tag{5.2.1}$$

$$(1-1.64B+.94B^2)(1-.18B-.2B^2-.1B^3-.26B^4-.15B^5-.16B^6)X_t$$
$$= (1-.58B)a_t \tag{5.2.2}$$

and

$$(1-1.66B+0.96B^2)X_t$$
$$= (1-.39B-0.09B^2-.09B^3+.25B^4+.11B^5+.18B^6)a_t . \tag{5.2.3}$$

Woodward and Gray compare these models to the well-known ARMA(2,0) model suggested by the Box and Jenkins method and claim that their models are better.

As mentioned in section 3.1, Wolfer's sunspot data present some kind of periodicity with unequal period from time to time; thus a model like (4.2.1) may be a good approach to it. In reviewing Fig. 2.2, we also notice that the local maximum of Wolfer's sunspot data fluctuate as a cosine function. So we substitute a_0 of model (4.2.1) with another jittery cosine wave to form a new model, and use the new model to fit the sunspot data. The model is defined as follows:

$$Y_t = \alpha+\beta(1+\gamma\cos(\omega_1 t+\theta_t^*))(\cos(\omega_0 t+\theta_t)+1) \qquad (5.2.4)$$

where

α, β and γ are some positive real constants,

$\theta_t = \sum_{i=0}^{t} \varepsilon_i$ is defined as in (4.2.1) and

$\theta_t^* = \sum_{i=0}^{t} \varepsilon_i^*$ with

$\varepsilon_0^* \sim U(0,2\pi)$, and

$\{\varepsilon_i^*, i\geq 1\}$ are i.i.d. normal random variables with zero mean and and variance σ_1^2, also all ε_i^*'s and ε_i's are mutually independent.

By following the steps of proof in theorem 4.2.1., we can prove that the process $\{Y_t\}$ defined in (5.2.4) is a stationary process too. The autocovariance of $\{Y_t\}$ is derived as follows:

$$E(Y_t) = \alpha+\beta \qquad t=1,2,\ldots$$

$$r(k) = EY_tY_{t+k}-(EY_t)^2$$

$$= \beta^2\{(\gamma^2 e^{-k\sigma_1^2/2}\cos(\omega_1 k))/2+(e^{-k\sigma^2/2}\cos(\omega_0 k))/2+$$

$$[\gamma^2 e^{-k(\sigma_1^2+\sigma^2)/2}\cos((\omega_1+\omega_0)k]/8+[\gamma^2 e^{-k(\sigma_1^2+\sigma^2)/2}\cos((\omega_1-\omega_0)k]/8\}$$

$$= \beta^2\{(\gamma^2 e^{-k\sigma_1^2/2}\cos(\omega_1 k)+e^{-k\sigma^2/2}\cos(\omega_0 k))/2+$$

$$(\gamma^2 e^{-k(\sigma^2+\sigma_1^2)/2}\cos(\omega_1 k)\cos(\omega_0 k))/4\} \qquad k=0,1,2,\ldots ;$$

thus

$$\mathrm{var}(Y_k) = r(0) = (1/2 +3\gamma^2/4)\beta^2$$

and the autocorrelation function is

$$\rho(k) = \{e^{-K\sigma^2/2}\cos(\omega_0 k)(1+\frac{\gamma^2}{2}e^{-K\sigma_1^2/2}\cos(\omega_1 k))/2+$$

$$(\gamma^2 e^{-K\sigma_1^2/2}\cos(\omega_1 k))/2\}/(1/2 +3\gamma^2/4) \quad . \qquad (5.2.5)$$

Although it is mentioned in section 4.4. that two different stationary processes may yield the same autocorrelation structure, comparing the sample autocorrelations from observed data with the theoretical autocorrelations of the model is one of the popular methods to check the adequacy of any proposed models. Thus, we fit the sample autocorrelations of sunspot data to equation (5.2.5) and obtain the estimated theoretical autocorrelations of model (5.2.4) for sunspot data. Since Woodward and Gray just give the plot of autocorrelations of their models, Procedure MATRIX of SAS package is used to produce these autocorrelations. The sample autocorrelations of sunspot data and the theoretical autocorrelations of model (5.2.1), model (5.2.2) and model (5.2.3) together with the estimated theoretical autocorrelations of model (5.2.4) are given in Table 5.1. The

Table 5.1

Autocorrelations of sunspot data.

K	$\rho_1(K)$	$\rho_2(K)$	$\rho_3(K)$	$\hat{\rho}_4(K)$	$\hat{\rho}$ (K)
1	0.82000	0.82680	0.79999	0.85974	0.81
2	0.34480	0.45660	0.38545	0.55218	0.43
3	-0.25453	0.03480	-0.05773	0.17673	0.03
4	-0.76223	-0.30690	-0.36961	-0.15482	-0.26
5	-0.99552	-0.50327	-0.50195	-0.36234	-0.40
6	-0.87043	-0.51216	-0.44856	-0.37233	-0.36
7	-0.43198	-0.34436	-0.26273	-0.22707	-0.17
8	0.16198	-0.04550	-0.00552	0.02355	0.10
9	0.69763	0.28499	0.24306	0.29349	0.34
10	0.98213	0.53938	0.40878	0.49799	0.49
11	0.91307	0.63784	0.44524	0.57826	0.50
12	0.51530	0.56097	0.34666	0.51649	0.37
13	-0.06797	0.34368	0.14803	0.33760	0.17
14	-0.62678	0.06021	-0.08706	0.09839	-0.04
15	-0.95994	-0.20371	-0.28663	-0.13153	-0.18
16	-0.94753	-0.37271	-0.39223	-0.29099	-0.25
17	-0.59400	-0.40315	-0.37594	-0.34408	-0.24
18	-0.02664	-0.29406	-0.24752	-0.28837	-0.19
19	0.55032	-0.08710	-0.04997	-0.15277	-0.10
20	0.92916	0.14857	0.15466	0.01342	0.01
21	0.97350	0.33903	0.30471	0.15625	0.12
22	0.66738	0.42897	0.35735	0.23304	0.20
23	0.12101	0.39690	0.30067	0.22363	0.20
24	-0.46893	0.25925	0.15606	0.13443	0.12
25	-0.89005	0.06289	-0.02958	-0.00546	-0.01
26	-0.99076	-0.13056	-0.19892	-0.15424	-0.16
27	-0.73479	-0.26390	-0.30181	-0.27028	-0.26
28	-0.21430	-0.30125	-0.31005	-0.32398	-0.30
29	0.38334	-0.23764	-0.22493	-0.30540	-0.30
30	0.84298	-0.09870	-0.07575	-0.22578	-0.26
31	0.99914	0.06883	0.09019	-0.11293	-0.16
32	0.79562	0.21251	0.22244	-0.00221	-0.02
33	0.30567	0.29026	0.28267	0.07403	0.09
34	-0.29432	0.28234	0.25568	0.09562	0.12
35	-0.78836	0.19590	0.15307	0.05956	0.07
36	-0.99858	0.06123	0.00865	-0.02004	-0.04
37	-0.84932	-0.07872	-0.13260	-0.11706	-0.15
38	-0.39430	-0.18195	-0.22841	-0.20152	-0.25
39	0.20267	-0.21998	-0.25187	-0.24854	-0.30
40	0.72667	-0.18556	-0.19883	-0.24526	-0.30
41	0.98908	-0.09365	-0.08826	-0.19402	-0.24
42	0.89541	0.02451	0.04436	-0.11090	-0.15
43	0.47940	0.13167	0.15837	-0.02041	-0.03

Table 5.1-continued.

K	$\rho_1(K)$	$\rho_2(K)$	$\rho_3(K)$	$\hat{\rho}_4(K)$	$\hat{\rho}(K)$
44	-0.10920	0.19612	0.22031	0.05209	0.07
45	-0.65848	0.20091	0.21368	0.08783	0.13
46	-0.97071	0.14798	0.14321	0.07976	0.13
47	-0.93349	0.05651	0.03259	0.03406	0.11
48	-0.56021	-0.04392	-0.08338	-0.03230	0.03
49	0.01475	-0.12279	-0.16969	-0.09711	-0.07
50	0.58440	-0.15787	-0.20165	-0.13967	-0.17
51	0.94366	-0.14141	-0.17183	-0.14688	-0.24
52	0.96321	-0.08157	-0.09166	-0.11667	-0.25
53	0.63600	0.00099	0.01281	-0.05817	-0.21
54	0.07983	0.08002	0.10925	0.01159	-0.11
55	-0.50508	0.13191	0.16906	0.07303	0.03

$\rho_1(K)$: Theoretical autocorrelations of model (5.2.1)

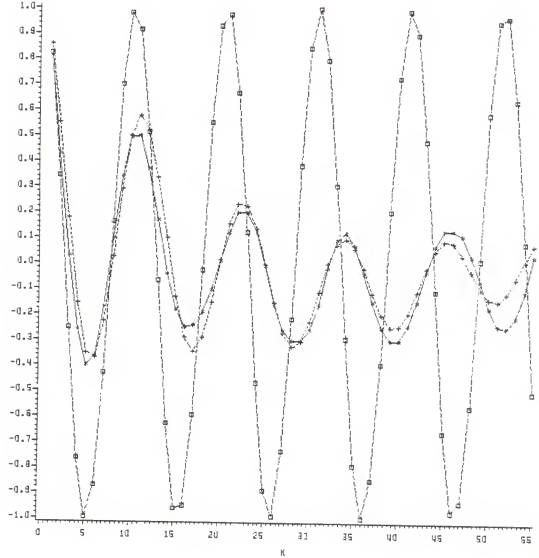
$\rho_2(K)$: Theoretical autocorrelations of model (5.2.2)

$\rho_3(K)$: Theoretical autocorrelations of model (5.2.3)

$\hat{\rho}_4(K)$: Estimated theoretical autocorrelations of model (5.2.4)

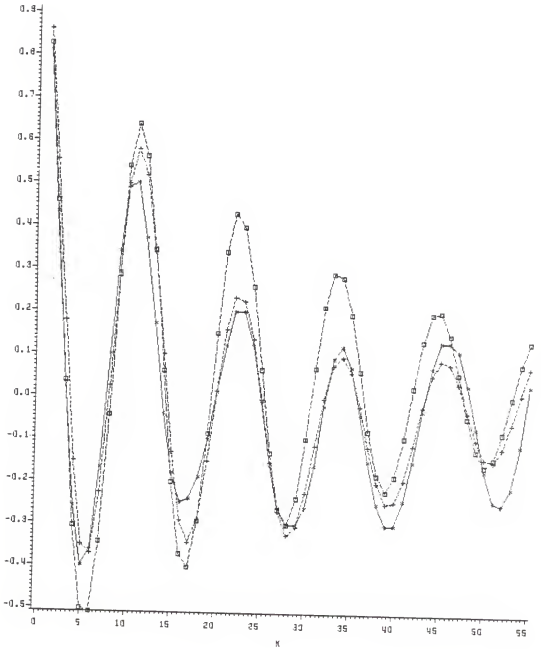
$\hat{\rho}(K)$: Sample autocorrelations from data

corresponding plots are shown respectively in Fig. 5.1, Fig. 5.2 and Fig. 5.3. These plots show that model (5.2.4) can fit sunspot data much better than model (5.2.1), model (5.2.2) and model (5.2.3)--at least in the sense of fitting autocorrelations. It would be necessary to develop a methodology to estimate the parameters of model (5.2.4) and predict the future values from this model. But we are unable to do it at this moment. This could be an interesting topic for future research.



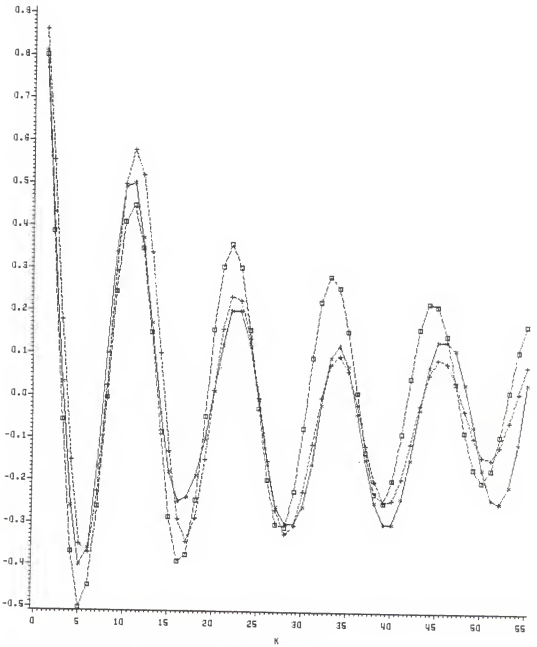
* denotes sample autocorrelations of data
□ denotes theoretical autocorrelations of model (5.2.1)
\+ denotes estimated theoretical autocorrelations of model (5.2.4)

Figure 5.1 Autocorrelations of sunspot data of model (5.2.1) and others.



* denotes sample autocorrelations of data
□ denotes theoretical autocorrelations of model (5.2.2)
\+ denotes estimated theoretical autocorrelations of model (5.2.4)

Figure 5.2 Autocorrelations of sunspot data of model (5.2.2) and others.



* denotes sample autocorrelations of data
□ denotes theoretical autocorrelations of model (5.2.3)
\+ denotes estimated theoretical autocorrelations of model (5.2.4)

Figure 5.3 Autocorrelations of sunspot data of model (5.2.3) and others.

BIBLIOGRAPHY

Anderson, O.D. (1975). Time Series Analysis and Forecasting. Butterworths, London.

Anderson, T.W. (1971). The Statistical Analysis of Time Series. John Wiley & Sons, Inc., New York.

Berman, S.M. (1964). "Limit Theorems for the Maximum Term in Stationary Sequences," Annals of Mathematical Statistics 35, 502-516.

Berman, S.M. (1971). "Asymptotic Independence of the Numbers of High and Low Level Crossings of Stationary Gaussian Process," Annals of Mathematical Statistics 42, 927-947.

Blake, I.F., and Lindsey, W.C. (1973). "Level-Crossing Problems for Random Processes," IEEE Transactions on Information Theory 19, 295-315.

Box, G.E.P., and Jenkins, G.M. (1976). Time Series Analysis, Forecasting, and Control. Holden Day, Inc., San Francisco.

Chatfield, C. (1980). The Analysis of Time Series: An Introduction. Chapman and Hall, New York.

Chow, Y.S., and Teicher, H. (1978). Probability Theory: Independence, Interchangeability, Martingales. Springer-Verlag New York, Inc., New York.

David, H.A. (1970). Order Statistics. John Wiley & Sons, Inc., New York.

Epstein, B. (1949). "The Distribution of Extreme Values in Samples Whose Members Are Subject to a Markoff Chain Condition," Annals of Mathematical Statistics 20, 590-594.

Epstein, B. (1951). "Correction to "The Distribution of Extreme Values in Samples Whose Members Are Subject to a Markoff Chain Condition"," Annals of Mathematical Statistics 22, 133-134.

Galambos, J. (1978). The Asymptotic Theory of Extreme Order Statistics. John Wiley & Sons, Inc., New York.

Graybill, F.A. (1969). Introduction to Matrices with Application in Statistics. Wadsworth Publishing Company, Inc., Belmont, California.

Gold, B. (1962). "Computer Program for Pitch Extraction," The Journal of the Acoustical Society of America 34(7), 916-921.

Gumbel, E.J. (1958). Statistics of Extremes. Columbia University Press, New York.

Gupta, S.S. (1963). "Probability Integrals of Multivariate Normal and Multivariate t," Annals of Mathematical Statistics 34, 792-828.

Hann, L. De. (1976). "Sample Extremes," Statistica Neerlandica 30, 161-172.

Horii, Y. (1975). "Some Statistical Characteristic of Voice Fundamental Frequency," Journal of Speech and Hearing Research 18, 192-201.

Horii, Y. (1979). "Fundamental Frequency Perturbation Observed in Sustained Phonation," Journal of Speech and Hearing Research 6, 5-19.

Hunter, D. (1977). "Approximating Percentage Points of Statistic Expressible as Maxima," TIMS Students in the Management Sciences 7, 25-36.

Leadbetter, M.R. (1974). "On Extreme Values in Stationary Sequences," Z. Wahrsh. Verw. Geb. 28, 288-303.

Leadbetter, M.R., Lindgren, G., and Rootzen, H. (1978). "Conditions for the Convergence in Distribution of Maxima of Stationary Normal Processes," Stochastic Processes and Their Applications 8, 131-139.

Leadbetter, M.R., Lindgren, G., and Rootzen, H. (1979). "Extreme and Related Properties of Stationary Processes. Part I: Extremes of Stationary Sequences," Statistical Research Report No. 1979-2, Department of Mathematics and Statistics, University of Unea, Sweden.

Moran, P.A.P. (1954). "Some Experiments on the Prediction of Sunspot Numbers," Journal of the Royal Statistical Society B16, 112-117.

Nelson, C.R. (1973). Applied Time Series Analysis for Managerial Forecasting. Holden-Day, Inc., San Francisco.

Slepian, D. (1962). "The One-sided Barrier Problem for Gaussian Noise," Bell System Technical Journal 41, 463-501.

Watson, G.S. (1954). "Extreme Values in Samples from M-Dependent Stationary Stochastic Processes," Annals of Mathematical Statistics 25, 788-800.

Woodward, W.A., and Gray, H.L. (1978). "New ARMA Models for Wolfer's Sunspot Data," Communications in Statistics, Part B: Simulation and Computation 7(1), 97-115.

Wu, C.F. (1981). "Asymptotic Theory of Nonlinear Least Squares Estimation," The Annals of Statistics 9(3), 501-513.

BIOGRAPHICAL SKETCH

Pan-Yu Lai was born on May 14, 1955, in Taipei, Taiwan, the Republic of China. He received a Bachelor of Science degree in applied mathematics in June, 1977, from National Chiao-Tung University in Taiwan.

Pan-Yu came to the University of FLorida in 1978 to pursue a master's degree in statistics. After he earned his degree in 1980, he decided to continue his education in statistics. While studying towards the degree of Doctor of Philosophy, Pan-Yu has worked as a graduate assistant performing statistical consulting duties in the Institute of Food and Agriculture Sciences among other places.

He is married to the former Hung-Ir Li of Taichung, Taiwan, and has a daughter, Shue-Fong.

On March 18, 1985, he will join the Statistics Unit of the Upjohn Company in Kalamazoo, Michigan, as a statistician.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mark C.K. Yang, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William E. Brownell
Associate Professor of Neuroscience

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Andrew J. Rosalsky
Associate Professor of Statistics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May, 1985

Dean for Graduate Studies and Research